UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea in Ingegneria Informatica e dell'Automazione

TESI DI LAUREA

Esperienze nel contesto dell'Intelligenza Artificiale: Riconoscimento di Immagini, Rilevamento delle Frodi e Information Retrieval

Experiences in the context of Artificial Intelligence: Image Recognition, Fraud Detection and Information Retrieval

Relatore

Prof. Domenico Ursino

Candidata

Valeria Cannone

ANNO ACCADEMICO 2023-2024

A chi sbaglia, ma impara dai propri errori.
A chi non sempre riesce a dare del proprio meglio.
A chi spesso non si sente all'altezza.
A chi è in ritardo, ma in realtà è nel suo tempo più giusto.
A chi, come me, ha deciso di non mollare, nonostante le difficoltà.
Datevi fiducia.
Sappiate che ce la farete, ma per farlo, abbiate cura di voi.

A mamma e papà,
Alle mie sorelle, Roberta e Benedetta,
pilastri fondamentali della mia vita.

Sommario

In questa tesi, analizzeremo in profondità l'applicazione delle tecnologie di Intelligenza Artificiale (IA) in contesti aziendali, evidenziando come le soluzioni basate su cloud possano trasformare le operazioni quotidiane delle organizzazioni. Ci concentreremo sull'uso di servizi cloud di Machine Learning, analizzando le funzionalità avanzate di riconoscimento di immagini, rilevamento delle frodi e Information Retrieval. Questa analisi ha l'obiettivo di dimostrare come tali soluzioni non solo migliorano l'efficienza operativa, ma anche ottimizzano la gestione delle informazioni, consentendo alle aziende di prendere decisioni più consapevoli e tempestive. Inoltre, attraverso casi studio e implementazioni pratiche, evidenzieremo i benefici concreti di queste tecnologie, sottolineando l'importanza di adottare un approccio strategico per integrare l'IA nei processi aziendali.

Keyword: Intelligenza Artficiale, Machine Learning, Cloud Computing, Amazon Web Services, Image Recognition, Information Retrieval, Rilevamento delle frodi, Elaborazione del Linguaggio Naturale (NLP).

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

Negli ultimi anni, l'Intelligenza Artificiale (IA) ha registrato una crescita esponenziale, diventando uno degli strumenti più discussi e ricercati nel panorama tecnologico globale. L'introduzione di modelli avanzati, come ChatGPT, ha sollevato interrogativi e critiche, in particolare riguardo all'impatto sul lavoro e sulla privacy, ma ha anche messo in evidenza il potenziale rivoluzionario di questa tecnologia.

L'IA simula i processi cognitivi umani, come l'apprendimento, il ragionamento e la pianificazione, il che ha portato alla sua diffusione in settori sempre più ampi, tra cui l'automazione industriale, la medicina, la finanza e il commercio elettronico.

Grazie ai progressi nel Machine Learning, nel Deep Learning e nell'elaborazione del linguaggio naturale (NLP), l'IA ha assunto un ruolo centrale nella trasformazione digitale delle aziende e delle istituzioni. Tuttavia, l'implementazione di soluzioni di Intelligenza Artificiale in ambienti complessi richiede infrastrutture di calcolo potenti e flessibili, in grado di adattarsi alle crescenti esigenze di dati e prestazioni. È in questo contesto che il Cloud Computing si rivela fondamentale, offrendo un accesso scalabile e flessibile a risorse computazionali avanzate. Amazon Web Services (AWS) si distingue come uno dei principali fornitori di servizi cloud, proponendo una vasta gamma di strumenti che consentono l'integrazione di funzionalità IA in diverse applicazioni.

La presente tesi si colloca all'interno di questo scenario di evoluzione tecnologica, con l'obiettivo di esplorare come l'Intelligenza Artificiale e il Cloud Computing, in particolare attraverso i servizi di AWS, possano migliorare l'efficienza operativa e ottimizzare la gestione dei dati. Nello specifico, abbiamo scelto di analizzare alcune delle soluzioni più innovative, come il riconoscimento di immagini, il rilevamento delle frodi e il recupero delle informazioni, con l'obiettivo di esaminare il loro impatto sui processi aziendali e di mettere in luce le opportunità derivanti dalla loro integrazione.

A questo punto, ci si potrebbe chiedere: Cosa dicono le statistiche riguardo all'evoluzione dell'Intelligenza Artificiale? Le prospettive per l'Intelligenza Artificiale appaiono certamente promettenti. Infatti, secondo il Grand View Research, si prevede un tasso di crescita annuale del 37,3% tra il 2023 e il 2030. La Cina, ad esempio, detiene già il primato mondiale per l'utilizzo di questa tecnologia, con un valore del 58%. Inoltre, secondo il World Economic Forum, l'IA dovrebbe generare circa 97 milioni di nuovi posti di lavoro.

In sintesi, sebbene l'IA possa sollevare preoccupazioni, le opportunità offerte per la creazione di nuovi posti di lavoro e la trasformazione dei settori economici sono significative. Il futuro dell'Intelligenza Artificiale dipenderà dalla nostra capacità di adottare queste tecnologie con responsabilità e visione a lungo termine, garantendo che il progresso tecnologico

proceda di pari passo con il benessere della società, al fine di trarre il massimo beneficio da esso.

La presente tesi è composta da sette capitoli strutturati come di seguito specificato:

- Nel Capitolo 1 sarà introdotto il concetto di Intelligenza Artificiale e, successivamente, studierà il funzionamento delle reti neurali e del Machine Learning.
- Nel Capitolo 2 si parlerà della piattaforma di Cloud Computing di Amazon, ovvero Amazon Web Service, approfondendo la sua struttura e i servizi che offre.
- Nel Capitolo 3 studieremo il riconoscimento di immagini e, in particolare, il servizio Amazon Rekognition. Successivamente, illustreremo come viene utilizzato tale servizio e quali funzionalità offre.
- Nel Capitolo 4 verranno analizzate le esperienze svolte sul rilevamento delle frodi con Amazon Fraud Detector; inoltre, osserveremo quali pratiche di sicurezza adottare per un sistema sicuro.
- Nel Capitolo 5 verrà illustrato il funzionamento dell'Information Retrieval con Amazon Kendra e il ruolo che svolge l'NLP in questa tecnologia.
- Nel Capitolo 6 saranno riportate le discussioni sui servizi di AWS utilizzati.
- Nel Capitolo 7 verranno tratte le conclusioni in merito al lavoro svolto e verranno evidenziati alcuni possibili sviluppi futuri nel contesto dell'Intelligenza Artificiale.

CAPITOLO 1

L'Intelligenza Artificiale

In questo capitolo verranno affrontati i diversi aspetti dell'Intelligenza Artificiale, dalla sua definizione ai suoi componenti, quali il Machine Learning, le reti neurali artificiali e il Deep Learning.

Si farà riferimento alle varie applicazioni in diversi ambiti della vita, come la medicina, il mondo digitale e l'informatica in generale. Verranno affrontati anche aspetti più complessi e matematici, soprattutto per quanto riguarda le reti neurali artificiali.

1.1 Definizione di Intelligenza Artificiale

La norma ISO/IEC 42001:2023 Information technology - Artificial Intelligence Management System (AIMS) definisce l'Intelligenza Artificiale come la capacità di un sistema di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività.

Il concetto di Intelligenza Artificiale (IA) ha molte sfumature. In passato, diversi ricercatori hanno definito questo strumento in modi differenti. Alcuni interpretano l'intelligenza come la capacità di ispirarsi al comportamento umano, mentre altri prediligono una definizione più concettuale e formale, conosciuta come razionalità, che in sostanza significa agire in modo corretto. Stessa situazione per quanto riguarda la concezione di razionalità, alcuni considerano l'intelligenza una proprietà dei processi di pensiero interni e ragionamento, mentre altri si concentrano sul comportamento intelligente.

1.2 Intelligenza Artificiale debole e forte

L'IA si distingue in:

- Intelligenza Artificiale debole o ristretta;
- Intelligenza Artificiale forte o generale.

Questa distinzione è dovuta alle diverse applicazioni e ai diversi approcci di questo potente strumento basati sulla mente umana.

1.2.1 Intelligenza Artificiale debole

L'Intelligenza Artificiale debole imita il comportamento umano applicando delle regole stabilite da un programmatore, il quale implementa funzioni predefinite per svolgere una

determinata azione. Quando si parla di IA debole possiamo pensare a strumenti come Siri, Alexa o la ricerca di Google. Infatti, questi, attraverso algoritmi elaborano i dati per trovare informazioni che conoscono e che classificheranno in base all'addestramento.

John Searle, in riferimento all'Intelligenza Artificiale debole, diceva: "*sarebbe utile per testare ipotesi sulle menti, ma non sarebbe in realtà una mente*", in quanto essa non ha lo scopo di possedere abilità cognitive generali, ma piuttosto di essere in grado di risolvere esattamente un singolo problema.

Nell'IA debole non esiste la necessità di comprendere totalmente i processi cognitivi dell'uomo. Essa si occupa esclusivamente del famoso problem solving, ovvero la risoluzione di problemi.

1.2.2 Intelligenza Artificiale forte

Al contrario dell'Intelligenza Artificiale debole, l'IA forte riesce a mettere in atto il pensiero umano ed è in grado di svolgere qualsiasi tipo di compito autonomamente risolvendo eventuali problemi o situazioni complicate attraverso una mente propria molto avanzata.

Il filosofo statunitense John Searle ha analizzato e contestato il concetto di Intelligenza Artificiale forte nel suo esperimento della stanza cinese.

Esperimento della stanza cinese di Searle e Test di Turing

Searle presentò l'argomentazione della *Stanza cinese* nel suo articolo *"Minds, Brains and Programs"*(Menti, cervelli e programmi) pubblicato nel 1980.

È stato elemento di dibattito per contestare il *Test di Turing* studiato da Alan Turing nel 1950. Turing, infatti, propose questo criterio per determinare se una macchina è in grado di esibire un comportamento intelligente indistinguibile da quello di un essere umano.

Il test era composto da 3 soggetti: un interrogatore umano, un uomo e una macchina.

Il primo soggetto aveva il compito di distinguere una risposta testuale umana da quella del computer. Se, dopo un periodo di tempo, l'esaminatore non riesce a distinguere con certezza quale delle due entità è la macchina e quale è l'essere umano, allora la macchina ha superato il test e quindi sarà dimostrato che la macchina possiede un'Intelligenza Artificiale, ovvero in questo caso, la capacità di imitare l'Intelligenza Artificiale umana in una conversazione scritta.

Dunque, John Searle con l'*esperimento della stanza cinese* contestò il criterio di Turing sostenendo che un computer non possa realmente "capire" o possedere una mente, anche se sembra comportarsi come se lo facesse.

L'esperimento spiega che, anche se in futuro si potesse costruire un computer che si comporta come se capisse il cinese, ciò non significherebbe che esso comprenda realmente la lingua. Basti immaginare un computer che prende simboli cinesi in ingresso e consulta una grande tabella per produrre altri simboli cinesi in uscita. Se il comportamento di questo computer fosse così convincente da superare il test di Turing, potrebbe convincere un cinese madrelingua che stesse parlando con un altro essere umano.

I sostenitori dell'Intelligenza Artificiale forte potrebbero concludere che il computer capisce il cinese, poiché il suo comportamento è indistinguibile da quello di un essere umano che conosce la lingua. Tuttavia, Searle chiede di immaginare che egli stesso si sieda all'interno del computer. In questa situazione, Searle si troverebbe in una piccola stanza (la stanza cinese), ricevendo simboli cinesi e utilizzando una tabella per produrre simboli cinesi in uscita. Nonostante ciò, Searle non capirebbe i simboli cinesi.

La mancanza di comprensione di Searle dimostra che il computer non può comprendere il cinese, poiché esso si trova nella stessa situazione. Il computer è un semplice manipolatore

di simboli, esattamente come lo è Searle nella stanza cinese. Quindi, anche se il computer risponde in modo coerente e appropriato, non capisce realmente ciò che sta dicendo.

Per questo il filosofo John Searle con il suo esperimento ha voluto dimostrare che il comportamento simile a quello umano non implica una vera comprensione o intelligenza. Anche se un computer può sembrare di capire una lingua, sta semplicemente seguendo delle regole predefinite senza alcuna comprensione del loro significato. Nella Figura 1.1 viene riassunto il rapporto esistente tra IA forte, IA debole e automazione.

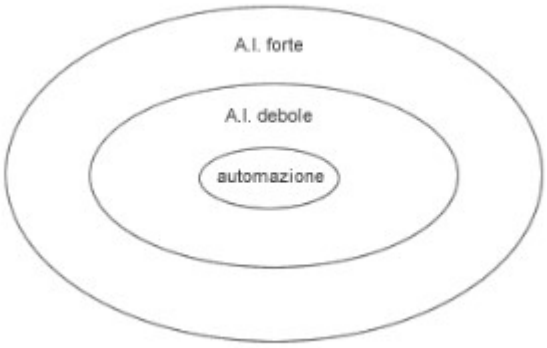


Figura 1.1: Funzionamento dell'IA debole e forte

1.3 Applicazioni dell'Intelligenza Artificiale

L'IA può essere applicata a diversi ambiti della vita reale. L'Osservatorio Artificial Intelligence ha distinto sei classi di applicazione di questo strumento in base alle diverse finalità di utilizzo. Infatti, proprio per questo motivo, l'IA sta diventando sempre di più uno strumento potente, capace di semplificare la vita delle persone.

L'IA viene applicata nel campo della medicina, nel campo finanziario, nel settore educativo, nei trasporti, nella logistica e nella produzione.

1.3.1 Medicina

In campo medico l'IA viene utilizzata soprattutto come supporto diagnostico, in particolare nell'area oncologica, respiratoria o cardiologica.

Come viene applicata però l'Intelligenza Artificiale in questo settore? Per prima cosa esso viene addestrato tramite radiografie, ecografie, TAC, elettrocardiogrammi e campioni di tessuti in modo da riuscire ad identificare patologie tumorali, cardiovascolari, dermatologiche e respiratorie con buona affidabilità.

Negli ultimi tempi si sta pensando anche ad addestrare modelli per fornire situazioni predittive utile a prevenire eventuali patologie ed essere pronti ad intervenire, così da facilitare e velocizzare il lavoro di medici e studiosi. Inoltre, questo modo di procedere è in grado di rilevare eventuali "pericoli" anche prima della manifestazione di sintomi, addirittura anche sei anni in anticipo.

Inoltre, l'IA è utilizzata per lo screening di molecole promettenti per ridurre i tempi di trasferimento dei risultati della ricerca alla pratica clinica.

Negli Stati Uniti, la *Food and Drug Administration* ha approvato oltre 500 applicazioni di IA, mentre in Italia la maggior parte è ancora in fase di sperimentazione.

Utilizzare l'Intelligenza Artificiale nella sanità non bisogna considerarlo come un problema, ma, al contrario, può essere utile per ridurre i tempi di ricerca e sperimentazione necessari ad effettuare diagnosi. Questo infatti non significa sostituire il lavoro di un medico e può considerarsi un aiuto molto valido.

Rischi

Uno dei principali rischi dell'uso dell'Intelligenza Artificiale in medicina è la mancanza di sufficienti test e supporto scientifico. È necessario condurre studi clinici più solidi e metodologicamente rigorosi, coinvolgendo più centri e valutando gli effetti in modo casuale su un campione rappresentativo per garantire l'affidabilità dei sistemi di IA.

Il documento "Linee guida sull'uso dei sistemi Intelligenza Artificiale in ambito diagnostico", di Eugenio Santoro e pubblicato dal Ministero della Salute, sottolinea l'importanza di questi studi. Inoltre, i sistemi di IA devono essere adeguatamente istruiti per evitare distorsioni, poiché vi sono casi documentati di fallimenti dovuti a campioni di addestramento non rappresentativi.

Un altro problema è il fenomeno della "black box", dove le risposte fornite dall'IA possono risultare incomprensibili. Tuttavia, l'IA non sostituirà i medici: le decisioni finali rimarranno agli specialisti per ragioni etiche, deontologiche e di responsabilità.

Infine, è necessaria una regolamentazione istituzionale, in linea con la nuova normativa europea sui dispositivi medici, per garantire la sicurezza e l'efficacia degli strumenti di IA prima della loro approvazione e commercializzazione.

1.3.2 IoT

IoT è l'acronimo di Internet of Things, ovvero l'Internet degli oggetti *intelligenti*. Si parla di oggetti smart quando facciamo riferimento a smartphone, computer o tablet, ma soprattutto si fa riferimento ad oggetti presenti all'interno delle case, sul luogo di lavoro o nelle città.

Alcuni esempi di applicazioni dell'IoT sono:

- la casa intelligente, o Smart Home;
- lo smart building, o edifici intelligenti;
- lo smart metering;
- la smart factory;
- auto intelligenti o Smart Car;
- le città intelligenti o Smart City.

In questo contesto, l'Intelligenza Artificiale viene applicata principalmente nelle Smart Factory, Smart Home e Smart City.

Negli ultimi tempi si parla sempre di più di Smart Home, ovvero *automatizzare* la casa permettendo alle persone di gestire e controllare la propria abitazione tramite dispositivi mobili. Tutto questo è possibile grazie ai sistemi di IA configurati su dispositivi come Amazon Echo che utilizza l'assistente personale intelligente Alexa, un elaboratore del linguaggio naturale.

Elaborazione del linguaggio naturale (NLP)

Alexa è un elaboratore del linguaggio naturale (NLP). Esso, attraverso il linguaggio parlato delle persone che lo utilizzano, è in grado di rispondere alle loro richieste. L'NLP è una vera e propria branca dell'Intelligenza Artificiale.

1.3.3 Chatbot

I chatbot sono sistemi di Intelligenza Artificiale usati maggiormente nel Customer Care aziendale per l'assistenza ai clienti. Essi sono in grado di eseguire azioni oppure offrire servizi ad un individuo sulla base di comandi vocali o testuali. La Generative IA, che contiene modelli linguistici elevati e soluzioni conversazionali, è in grado di rendere le risposte offerte molto più comprensibili e flessibili per facilitare la capacità di intendere dell'utente.

Esistono due tipi di Chatbot, ovvero:

- chatbot conversazionali;
- chatbot basati su regole.

Un esempio recente di IA Chatbot conversazionale è, sicuramente, ChatGPT, un'applicazione di Intelligenza Artificiale Generativa sviluppata da OpenAI per creare risposte che siano coerenti e pertinenti al contesto della conversazione testuale basata sul linguaggio naturale. Esempi di chatbot basati su regole sono i bot, che a differenza dei primi, dialogano con gli utenti fornendo risposte standard a domande standard.

1.3.4 Computer Vision

La Computer Vision studia algoritmi e tecniche per permettere ai computer di riprodurre funzioni e processi dell'apparato visivo umano. Ha il compito di riconoscere e comprendere ogni singolo elemento all'interno di immagini singole o video al fine di estrarre grandi quantità di informazioni di diverso tipo; le più frequenti sono quelle biometriche.

Per poter funzionare correttamente, i sistemi di computer vision devono essere addestrati con una quantità notevole di immagini opportunamente etichettate.

I principali task di computer vision sono i seguenti:

- *Image Classification:* analisi del contenuto delle immagini e rilevamento delle etichette;
- *Object Detection:* rilevamento dell'entità;
- *Image Segmentation:* suddivisione delle immagini in sezioni;
- *Face Recognition:* riconoscimento di volti;
- *Action Recognition:* identificazione di una o più entità nello svolgimento di una determinata azione;
- *Visual Relationship Detection:* comprensione delle relazioni tra gli oggetti in un'immagine;
- *Emotion Recognition:* rilevamento del sentiment;
- *Image Editing:*modifica di un'immagine per l'oscuramento di dati sensibili.

1.3.5 Intelligent Data Processing (IDP)

L'IDP è una tecnologia utilizzata per l'estrazione di diverse tipologie di informazioni presenti all'interno di enormi quantità di dati. L'obiettivo di questa tecnologia è velocizzare il lavoro delle persone e ridurre i costi per il personale.

Esso viene usato, ad esempio, all'interno delle aziende per classificare i dati oppure per ricercare facilmente quello di cui hanno bisogno.

In questa categoria, rientrano in particolare, l'analisi predittiva e il rilevamento di frodi.

I settori che utilizzano questa tecnica, sono, infatti, quelli bancari e finanziari, sanitari, legali, i siti di e-commerce e la Pubblica Amministrazione.

1.3.6 Recommender System

I sistemi di raccomandazione (Recommender System) sono un sottocampo dell'Intelligenza Artificiale e del Machine Learning che si occupa di suggerire agli utenti contenuti, prodotti o servizi che potrebbero trovare interessanti. Sono utilizzati in una vasta gamma di applicazioni, dai motori di ricerca alle piattaforme di e-commerce, dai servizi di streaming ai social network.

Tipologie di Recommender System

I Recommender System si suddividono nelle seguenti categorie:

- *Collaborative Filtering (Filtraggio Collaborativo)*, questa categoria, a sua volta, si suddivide in:
  - *User-based Collaborative Filtering:* raccomanda articoli che utenti simili hanno apprezzato. Si basa sull'assunto che se un utente A ha gusti simili a un utente B, i contenuti apprezzati da B saranno apprezzati anche da A.
  - *Item-based Collaborative Filtering:* raccomanda articoli simili a quelli che l'utente ha apprezzato in passato. Si basa sulla similarità tra articoli, suggerendo contenuti che hanno ricevuto valutazioni simili da parte di altri utenti.
- *Content-based Filtering (Filtraggio Basato sul Contenuto):* raccomanda articoli simili a quelli che l'utente ha apprezzato, basandosi sulle caratteristiche del contenuto stesso (ad esempio, generi di film, caratteristiche dei prodotti).
- *Hybrid Systems (Sistemi Ibridi):* combinano più tecniche di raccomandazione per migliorare la precisione delle raccomandazioni. Ad esempio, possono integrare filtraggio collaborativo e filtraggio basato sul contenuto per superare le limitazioni di ciascun metodo.

1.4 Reti neurali

Le reti neurali sono una particolare tecnica di Machine Learning e sono la base di avanzate forme di Intelligenza Artificiale. Esse tendono a simulare l'apprendimento di un cervello umano facendo riferimento al suo meccanismo naturale.

Anatomicamente, infatti, il cervello è composto da cellule, dette neuroni, che sono collegate tra loro tramite assoni e dendriti formando le sinapsi. Questa connessione spesso cambia in risposta agli stimoli esterni ed è proprio questo cambiamento che viene simulato tramite le reti neurali artificiali.

Nel caso delle reti neurali artificiali, gli stimoli esterni derivano dai dati forniti per l'addestramento, formati da una componente di input e una componente di output. Ad esempio, se viene fornita un'immagine in ingresso, la rete neurale restituirà un'etichetta esplicativa associata.

Struttura

Una rete neurale è rappresentata dal grafo di flusso. Esso è composto da nodi, collegati tra loro attraverso gli archi. I nodi indicano le unità in cui vengono effettuate determinate operazioni, sia semplici che più complesse, da una semplice operazione matematica (addizione, sottrazione, moltiplicazione) all'esecuzione di algoritmi complessi.

In un grafo, i nodi si dividono in 3 tipologie:

- *nodi di entrata (input node):* utilizzati per introdurre dati in input (Figura 1.2);
- *nodi intermedi (hidden node):* strato in cui avviene l'elaborazione dei dati (Figura 1.3);
- *nodi di uscita (output node):* restituiscono il risultato dell'elaborazione (Figura 1.4).

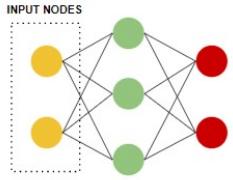


Figura 1.2: Nodi di input

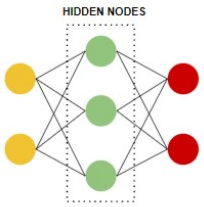


Figura 1.3: Nodi nascosti

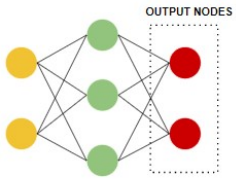


Figura 1.4: Nodi di output

Ogni nodo è collegato ad altri nodi e possiede un proprio peso e una soglia specifici. Se l'output di un nodo supera la soglia definita, il nodo si attiva e trasmette i dati al livello successivo della rete; invece, se l'output non raggiunge la soglia, il nodo non invia alcun dato al livello successivo.

Gli archi trasmettono i dati da un nodo all'altro e possono collegare i nodi sia in entrata che in uscita. Ogni arco possiede un peso sinaptico, rappresentato da un coefficiente C, che permette di modificare il dato prima di mandarlo nel nodo di output.

La propagazione dell'informazione attraverso gli archi può avvenire in due modi differenti: in avanti o all'indietro.

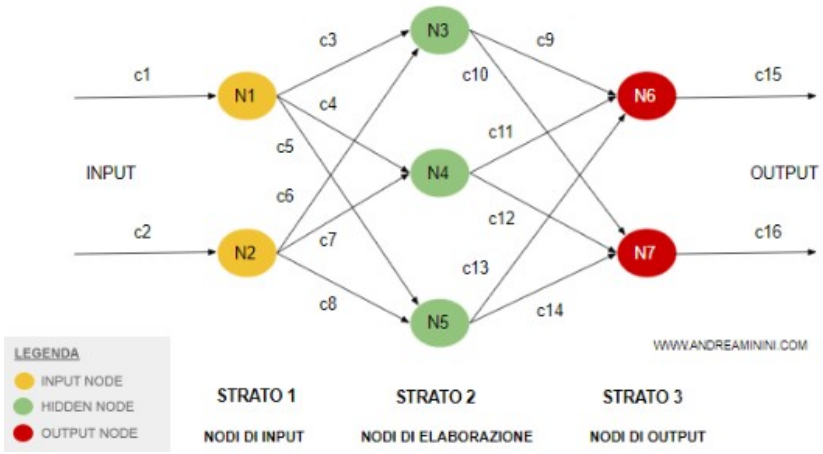


Figura 1.5: Rete neurale semplice

1.4.1 Propagazione Feed Forward

Nella *propagazione in avanti*, anche detta *Feed-Forward*, la trasmissione dell'informazione avviene da sinistra verso destra e segue una sola direzione, senza retroazione. Ogni nodo di questa configurazione si attiva solo se il suo output supera una soglia predefinita (Figura 1.5).

Si tratta di una configurazione semplice perché non crea cicli tra gli strati (Figura 1.6).

L'errore viene calcolato confrontando l'output della rete con l'output desiderato (etichetta) e la differenza ottenuta viene elaborata tramite due funzioni di perdita principali: errore quadratico medio (MSE) ed entropia incrociata.

Errore quadratico medio (MSE)

L'errore quadratico medio è una misura della differenza media al quadrato tra i valori previsti dalla rete neurale (y*) e i valori reali (y).

Entropia incrociata

L'entropia incrociata misura la differenza tra due distribuzioni di probabilità, ovvero la distribuzione delle etichette reali e quella delle etichette previste. È spesso utilizzata nei problemi di classificazione.

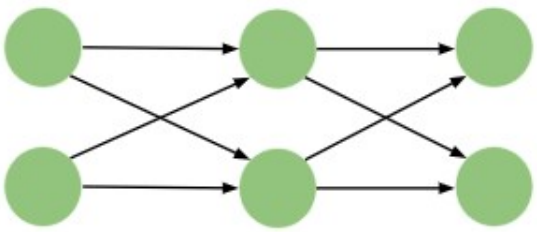


Figura 1.6: Rete neurale feed forward

1.4.2 Propagazione Feed-Back

Nella *propagazione all'indietro*, detta *Feed-Back* (Figura 1.7), il risultato viene trasmesso nel primo nodo a sinistra. Essa consente di aggiornare i pesi e i bias dei nodi in modo da minimizzare l'errore tra l'output previsto dalla rete e l'output desiderato. Le reti possono diventare anche cicliche.

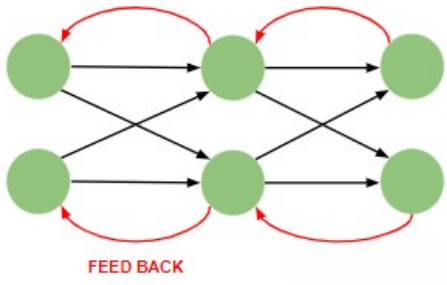


Figura 1.7: Rete neurale feed-back a retroazione

1.4.3 Architettura reti neurali

La reti neurali possono dividersi, in base alla loro architettura, in:

- single-layer;
- multi-layer.

1.4.4 Reti single-layer

Il primo schema di rete neurale a singolo livello è stato introdotto da Frank Rosenblatt, nella rivista Psychological review nell'anno 1958. Questo schema si chiama Perceptron (percettore) e contiene un singolo livello di input e un nodo di output.

Il modello probabilistico di Rosenblatt si concentra sull'analisi matematica di funzioni come l'immagazzinamento delle informazioni e la loro influenza nel riconoscimento dei pattern. Esso rappresenta un progresso significativo rispetto al modello binario di McCulloch e Pitts, poiché i pesi sinaptici nel percettrone di Rosenblatt sono variabili, permettendo, così, al sistema di apprendere.

Ogni istanza di training ha la forma (X, y), dove X è un vettore di variabili di funzionalità e y è la variabile di classe binaria (-1 o $+1$). L'obiettivo è prevedere y per nuovi dati non etichettati. Nel percettrone, il livello di input trasmette le caratteristiche X al nodo di output senza eseguire calcoli. Il nodo di output calcola una funzione lineare $W * X$, dove W è il vettore dei pesi. La previsione y* è il segno di questa somma pesata: y* = segno($W * X$). La funzione del segno mappa il risultato su $+1$ o -1, adatta per la classificazione binaria. L'errore di previsione è E(X) = y − y*, che può essere -2, 0 o $+2$. Se l'errore è diverso da zero, i pesi vengono aggiornati nella direzione opposta al gradiente dell'errore. Ciò è simile all'aggiornamento dei pesi nei modelli di Machine Learning lineare.

Il percettrone è interpretato come un'unità computazionale, utile per combinare più unità e creare modelli complessi. Nonostante abbia due strati (input e output), il livello di input non esegue calcoli e, quindi, non viene contato, rendendo il percettrone una rete a singolo strato.

In contesti con distribuzioni di classe sbilanciate, è necessario aggiungere un bias b per catturare una parte invariante della previsione. Ciò permette di migliorare la previsione quando la media della classe binaria non è zero, bilanciando meglio il modello.

La rete Perceptron è detta anche rete feed-forward. Il processo di apprendimento è supervisionato e la rete è in grado di risolvere operazioni logiche di base come AND e OR.

Questa rete è utilizzata anche per la classificazione di modelli (pattern classification).

1.4.5 Multi-layer

A differenza delle reti single-layer, le multistrato (o multi-layer) contengono più strati computazionali.

Nelle reti multistrato, o MultiLayer Perceptron (MLP), le unità di elaborazione sono di tre tipi: Input (ricevono input dall'esterno), Hidden (ricevono input dalle unità di ingresso) e Output (ricevono input dalle unità nascoste e trasmettono segnali all'esterno del sistema).

L'attività delle unità hidden non è visibile all'esterno, mentre l'elaborazione è frutto dello scambio di messaggi e informazioni fra neuroni (nodi), attraverso le connessioni (archi). Se la rete MLP ha più di due layer nascosti può essere definita *deep*, mentre normalmente viene detta *shallow*.

La tipologia di rete più comune è quella cosiddetta *densa (fully connected)*. Ogni neurone del layer $k-1$ è connesso con ciascun neurone del layer k.

La rete viene caratterizzata da una matrice di pesi (Figura 1.8) (più il vettore dei bias) per ciascun layer.

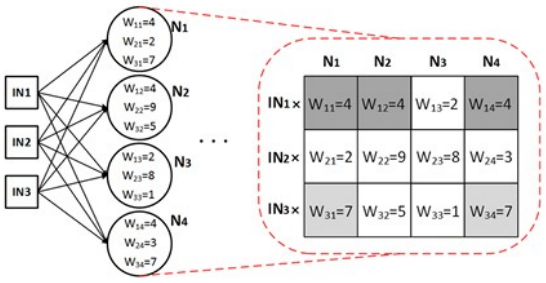


Figura 1.8: Matrice dei pesi

1.4.6 Valutazioni

Le reti neurali offrono numerosi vantaggi che le rendono strumenti potenti per una vasta gamma di applicazioni. Tra i principali vantaggi troviamo:

- *Capacità di classificare pattern complessi e non lineari:* le reti neurali eccellono nel riconoscere e classificare pattern intricati in dati come immagini, video, suoni e testi, grazie alla loro struttura multilivello e alle funzioni di attivazione non lineari.
- *Capacità di lavorare in parallelo:* le reti neurali possono eseguire elaborazioni parallele, sfruttando appieno le architetture hardware moderne, come GPU e TPU, il che le rende altamente efficienti nel trattamento di grandi quantità di dati.
- *Tolleranza agli errori e al rumore:* esse sono robuste rispetto a rumori ed errori nei dati di input, potendo ancora produrre risultati accurati anche quando i dati non sono perfettamente puliti.
- *Alta precisione:* esse possono raggiungere livelli elevati di accuratezza, soprattutto in compiti di classificazione e predizione, grazie alla loro capacità di apprendere da grandi volumi di dati.
- *Facilità di aggiornamento con nuovi dati:* le reti neurali possono essere facilmente riaddestrate con nuovi dati, permettendo l'aggiornamento continuo dei modelli senza necessità di riprogettare l'architettura di base.
- *Capacità di generalizzazione:* una volta addestrate, esse possono generalizzare bene su nuovi dati, ovvero riconoscere pattern non visti durante la fase di training.
- *Indipendenza da assunzioni a priori:* le reti neurali non richiedono ipotesi preliminari sui dati, come la distribuzione statistica o la linearità delle relazioni, il che le rende versatili in molteplici contesti.
- *Capacità di predire sia su problemi di regressione che di classificazione:* esse possono essere utilizzate efficacemente sia per problemi di predizione continua (regressione) che per problemi di classificazione.

Tuttavia, le reti neurali presentano anche alcuni svantaggi:

- *Problemi specifici:* esse sono progettate per risolvere problemi specifici e possono non essere la scelta migliore per tutte le tipologie di problemi.
- *Difficoltà nel trattare variabili categoriche con molti valori diversi:* per affrontare questo problema, è spesso necessario normalizzare i dataset, il che può essere complicato e dispendioso in termini di tempo.
- *Necessità di grandi dataset:* per ottenere buoni risultati, le reti neurali richiedono dataset di dimensioni considerevoli, il che può essere un limite in contesti dove i dati sono scarsi o difficili da raccogliere.
- *Tempi di training lunghi:* esse possono richiedere tempi di addestramento molto lunghi, soprattutto quando si utilizzano modelli complessi e dataset di grandi dimensioni.
- *Necessità di testare diverse configurazioni:* è spesso necessario testare e confrontare numerose configurazioni della rete per trovare quella ottimale, il che può essere un processo lungo e complesso.

- *Poca trasparenza del processamento dei dati:* le reti neurali sono spesso considerate "black box" a causa della difficoltà nell'interpretare e comprendere esattamente come vengono elaborati i dati al loro interno.
- *Computazionalmente onerose:* esse richiedono una potenza di calcolo elevata, il che può rendere oneroso il loro utilizzo, soprattutto in ambienti con risorse limitate.

1.5 Machine Learning

Il Machine Learning è una branca dell'Intelligenza Artificiale che si occupa di apprendimento automatico e si focalizza sullo sviluppo di algoritmi e modelli capaci di apprendere e migliorare le proprie prestazioni attraverso l'analisi dei dati.

Gli algoritmi usati principalmente per l'apprendimento automatico nel Machine Learning sono:

- apprendimento supervisionato;
- apprendimento non supervisionato;
- apprendimento per rinforzo.

1.5.1 Apprendimento supervisionato

È l'algoritmo più utilizzato nel Machine Learning. I modelli sono addestrati utilizzando dati etichettati, cioè dati per i quali conosciamo già la risposta. L'obiettivo è quello di costruire un modello che possa fare previsioni accurate su nuovi dati basandosi su queste etichette (Figura 1.9).

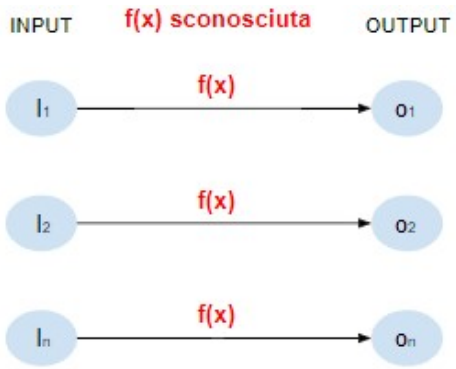


Figura 1.9: Apprendimento supervisionato

Gli algoritmi più comuni per questo tipo di apprendimento sono i seguenti:

- Regressione Lineare e Logistica;
- Alberi di Decisione e Foreste Casuali;
- Support Vector Machines (SVM);
- Reti Neurali.

1.5.2 Apprendimento non supervisionato

Nell'apprendimento non supervisionato, i modelli lavorano con dati non etichettati e cercano di identificare pattern, strutture o relazioni nascoste senza alcuna guida esterna (Figura 1.10).

Gli algoritmi più comuni per questo tipo di apprendimento sono i seguenti:

- K-means Clustering;
- Algoritmi di Clustering Gerarchico;
- Analisi delle Componenti Principali (PCA);
- T-Distributed Stochastic Neighbor Embedding (t-SNE).

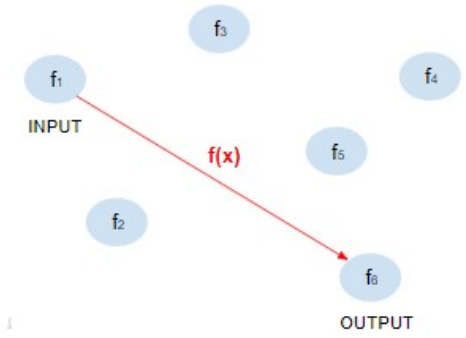


Figura 1.10: Apprendimento non supervisionato

1.5.3 Apprendimento per rinforzo

Nell'apprendimento per rinforzo, un agente interagisce con un ambiente e apprende a prendere decisioni basate sulle ricompense o penalità ricevute per le sue azioni. Esso non possiede esempi predefiniti di input e output o indicazioni su quali comportamenti seguire.

Tuttavia, è dotato di una funzione di rinforzo legata a un obiettivo specifico, che gli consente di valutare il feedback delle proprie azioni. Attraverso l'esperienza, l'agente impara a sviluppare una strategia comportamentale che massimizza il rinforzo, ovvero il premio. In altre parole, la macchina conosce l'obiettivo finale da raggiungere, ma non sa in anticipo quale percorso seguire per arrivarci.

L'obiettivo è massimizzare la ricompensa totale nel tempo (Figura 1.11).

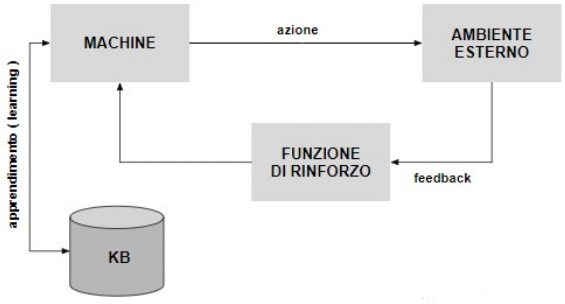


Figura 1.11: Apprendimento per rinforzo

1.5.4 Fasi del Machine Learning

Le fasi del Machine Learning sono le seguenti:

1. *Pre-processing:* questo è un passo in cui vengono analizzati i dati nel dataset di apprendimento per ottimizzare le performance dell'algoritmo. Vengono individuate le informazioni ridondanti e i rumori (dati irrilevanti) permettendo la riduzione delle dimensioni del dataset.

2. *Apprendimento:* questo è un passo in cui l'algoritmo apprende automaticamente dai dati presenti nel dataset di apprendimento (training set) per elaborare un modello di previsione.

3. *Valutazione:* questa è la fase di valutazione del modello di previsione creato dall'algoritmo per vedere se è valido o meno. Per valutare la qualità predittiva si fissa una percentuale T di risposte esatte come parametro di riferimento. Successivamente si prova il modello predittivo con un test set, diverso dal dataset di apprendimento, e si analizzano i risultati.

   Se $R \geq T$, ovvero la percentuale di risposte esatte R supera la soglia T, il modello predittivo è valido. Le risposte della macchina sono sufficientemente accurate e il margine di errore delle risposte è accettabile. Se $R < T$, la percentuale di risposte esatte R non supera la soglia T, il processo riparte dalla fase pre-processing perché la qualità previsionale non è accettabile.

4. *Predizione:* è la fase finale di applicazione, quella in cui il modello predittivo è usato con i dati reali per risolvere dei problemi pratici, da parte della macchina stessa o degli utenti finali.

1.5.5 Tipologie di Machine Learning

Deep Learning

Il Deep Learning è una sottocategoria dell'apprendimento automatico che utilizza reti neurali artificiali con molteplici strati (deep neural network) per modellare e risolvere problemi complessi. È particolarmente efficace nel trattare dati non strutturati come immagini, audio e testo. Grazie alla sua capacità di apprendere rappresentazioni gerarchiche dei dati, il Deep Learning ha rivoluzionato campi come la visione artificiale, il riconoscimento del linguaggio e molte altre applicazioni.

Real Time Machine Learning

L'apprendimento automatico in tempo reale è una branca dell'apprendimento automatico che si occupa di processare e analizzare i dati appena vengono raccolti per prendere decisioni immediate o quasi immediate. Questo approccio è particolarmente utile in applicazioni dove le decisioni devono essere prese rapidamente e i dati sono continuamente in arrivo.

Model Prediction

La predizione dei modelli di machine learning si riferisce al processo con cui un modello, addestrato su un set di dati, viene utilizzato per fare previsioni o classificazioni su nuovi dati. Tale processo è cruciale per l'applicazione pratica dei modelli di machine learning in vari contesti, come la diagnosi medica, il trading finanziario, le raccomandazioni personalizzate e molto altro.

Information Retrieval (IR)

Il recupero delle informazioni (Information Retrieval, IR) è il processo di ricerca e recupero di informazioni pertinenti da grandi collezioni di dati, spesso non strutturati, come documenti di testo, immagini, audio e video. L'obiettivo principale è quello di soddisfare una query dell'utente nel modo più efficace possibile, fornendo risultati rilevanti e precisi.

1.5.6 Processo del Machine Learning

I passaggi principali del processo di sviluppo di un modello di Machine Learning sono i seguenti:

- *Raccolta dei Dati:* vengono collezionati i dati di interesse.
- *Preprocessing dei Dati:* si procede con la pulizia dei dati, la gestione dei dati mancanti e la normalizzazione/scalatura.
- *Divisione dei Dati:* suddivisione in set di training, validazione e test.
- *Selezione del Modello:* si sceglie l'algoritmo appropriato per il problema specifico.
- *Addestramento del Modello:* si utilizzano i dati di training per addestrare il modello.
- *Valutazione del Modello:* si utilizzano i dati di validazione e test per valutare le prestazioni.
- *Tuning degli Iperparametri:* si ottimizzano i parametri del modello per migliorare le prestazioni.
- *Implementazione e Manutenzione:* si effettua il deployment del modello in un ambiente di produzione e monitoraggio delle sue prestazioni.

CAPITOLO 2

Amazon Web Services

In questo capitolo si parlerà di Amazon Web Services (AWS), una piattaforma leader nel cloud computing fornita da Amazon. Il capitolo esplora i principali vantaggi di AWS, l'importanza del cloud computing e come AWS fornisca servizi IT on-demand attraverso modelli di servizio come Infrastructure as a Service (IaaS), Platform as a Service (PaaS) e Software as a Service (SaaS). Il capitolo evidenzia come AWS supporti le aziende nel ridurre i costi IT, migliorare l'efficienza operativa e accelerare il time-to-market attraverso l'accesso a tecnologie avanzate. Esso infine, analizza i principali servizi della piattaforma cloud.

2.1 Introduzione ad AWS

Amazon Web Services, Inc. (AWS) è una delle principali piattaforme di cloud computing al mondo. Essa fornisce diversi servizi utilizzabili in 26 regioni geografiche suddivise secondo lo schema stabilito da Amazon.

In totale AWS offre 200 prodotti, utili per le aziende e gli individui, a creare ed eseguire le proprie applicazioni e i propri siti web secondo un modello Pay-as-you-Go, il che significa che vengono pagate solo le risorse utilizzate.

I prodotti offerti da AWS si dividono in diverse categorie, quali:

- servizi di calcolo;
- servizi di rete;
- servizi di distribuzione dei contenuti;
- servizi di archiviazione (storage);
- servizi di database;
- strumenti per sviluppatori;
- servizi di sicurezza;
- servizi di analisi;
- servizi di deployment;
- servizi di Machine Learning;
- Servizi di IoT.

In particolare, in riferimento ai servizi di Machine Learning, ne verranno analizzati alcuni nei capitoli successivi. In dettaglio, essi saranno Amazon Rekognition, Amazon Fraud Detector e Amazon Kendra.

2.1.1 Storia

Nel 2001, AWS nasce come una serie di servizi web per sviluppatori, offrendo API (Application Programming Interface) per l'integrazione con le applicazioni web. Questo rappresenta il primo tentativo di Amazon di offrire un'infrastruttura IT come servizio.

Nell'anno successivo, Jeff Bezos, fondatore di Amazon, e il suo team riconobbero l'esigenza di sviluppare una piattaforma cloud in grado di supportare la crescita e la scalabilità delle operazioni interne dell'azienda. Questa nuova infrastruttura avrebbe dovuto essere sufficientemente versatile e potente per gestire l'espansione continua e le crescenti esigenze operative di Amazon. Da questa intuizione nasce l'idea di offrire infrastruttura IT come servizio, accessibile via Internet.

Il primo servizio di AWS è stato introdotto nel 2004 e permetteva la gestione delle code di messaggi tra diversi sistemi distribuiti. Il suo nome era Amazon SQS (Simple Queue Service).

Con il passare degli anni però, Amazon, ha deciso di lanciare ufficialmente il suo sistema di cloud computing aggiungendo alla piattaforma diversi servizi, di cui due principali:

- *Amazon S3 (Simple Storage Service):* esso si occupa di archiviazione scalabile di oggetti nel cloud;
- *Amazon EC2 (Elastic Compute Cloud):* esso fornisce server virtuali scalabili su richiesta.

Questi servizi permettono alle aziende di utilizzare risorse IT, come ad esempio hardware e sistemi software, in modo flessibile e scalabile, pagando solo per l'uso effettivo.

Dal 2010, AWS espande la propria infrastruttura con data center in tutto il mondo, rendendo i suoi servizi disponibili a livello globale. Per questo motivo, viene introdotto Amazon VPC (Virtual Private Cloud) utilizzato per isolare le risorse in una rete virtuale.

Con l'emergere di nuove tecnologie e l'intensificarsi della concorrenza nel mercato, nel 2013, Amazon Web Service decise di innovare e ampliare la sua offerta con nuovi servizi, rispondendo così alle esigenze in continua evoluzione dei suoi clienti. Alcuni di questi sono:

- *Amazon Redshift (2013):* servizio di data warehousing veloce e scalabile;
- *AWS Kinesis (2013):* servizio di elaborazione in tempo reale di flussi di dati (ad esempio, registrazione di una videocamera di sorveglianza);
- *AWS Lambda (2014):* servizio che si occupa dell'esecuzione di codice senza provisioning di server.

La sua popolarità dipende soprattutto, però, dall'introduzione del Machine Learning, delle tecnologie IoT e del calcolo quantistico che l'hanno reso leader indiscusso nel mercato del cloud computing.

Negli ultimi tempi, in particolare dal 2021 ad oggi, AWS introduce nuove soluzioni sostenibili per ridurre l'impatto ambientale dei data center AWS. Esse sono:

- *AWS Local Zones:* espansione dei data center in regioni specifiche per bassa latenza;
- *AWS Graviton:* processori basati su ARM per prestazioni migliori e costi inferiori.

Amazon Web Service ha trasformato il modo in cui le aziende pensano all'infrastruttura IT, offrendo una piattaforma scalabile, flessibile e sicura che ridefinisce i limiti dell'innovazione tecnologica.

2.1.2 Architettura di AWS

L'architettura di AWS si basa su diverse componenti chiave che lavorano insieme per garantire prestazioni elevate e resilienza.

Essa è strutturata in regioni geografiche, ognuna delle quali comprende diverse *Availability Zone (AZ)* (Figura 2.1). Le regioni rappresentano aree geografiche distinte, come "US East" o "EU Central", mentre le AZ sono data center indipendenti all'interno di una regione.

Figura 2.1: Availability Zone di AWS

Questa struttura permette di progettare applicazioni ad alta disponibilità e con tolleranza ai guasti, distribuendo i carichi di lavoro su più AZ per garantire continuità operativa anche in caso di guasti localizzati.

Un altro elemento cruciale dell'architettura di Amazon Web Service è il Virtual Private Cloud (VPC). Amazon VPC permette ai clienti di creare reti virtuali isolate all'interno del cloud AWS, offrendo un controllo completo sulle configurazioni di rete. Con il VPC è possibile definire subnet, tabelle di routing e gateway, migliorando, così, la sicurezza e la gestione delle risorse. Questo significa che le aziende possono configurare la propria rete nel cloud in modo simile a come farebbero on-premises, ovvero in locale, con l'aggiunta della flessibilità e scalabilità del cloud.

L'elasticità è un principio fondamentale di AWS, realizzato attraverso servizi come Auto Scaling e Elastic Load Balancing (ELB). Auto Scaling consente di aggiungere o rimuovere automaticamente risorse in risposta alle variazioni della domanda, assicurando che le applicazioni rimangano performanti durante i picchi di traffico ed efficienti rispetto ai costi nei periodi di bassa attività. Elastic Load Balancing, invece, distribuisce il traffico in entrata su più istanze EC2, migliorando la disponibilità e la resilienza delle applicazioni. In questo modo, le aziende possono adattarsi rapidamente alle mutevoli esigenze del mercato senza preoccuparsi delle limitazioni infrastrutturali.

AWS adotta un approccio a più livelli per la sicurezza, includendo protezioni a livello di rete, di infrastruttura, di applicazione e dati. Il servizio Identity and Access Management (IAM) gestisce in modo granulare gli accessi e i permessi per le risorse AWS, garantendo che solo le persone autorizzate possano accedere a risorse sensibili. Inoltre, AWS offre servizi

come AWS Shield e Web Application Firewall (WAF) per proteggere il sistema dalle minacce esterne, come gli attacchi DDoS.

In aggiunta, AWS aderisce a numerosi standard di conformità, assicurando che le risorse siano protette e conformi alle normative vigenti.

Per quanto riguarda il monitoraggio e la gestione delle risorse, AWS offre strumenti come AWS CloudWatch e AWS CloudTrail. AWS CloudWatch monitora le risorse e le applicazioni in esecuzione su AWS, raccogliendo metriche, monitorando i log e impostando allarmi per rilevare anomalie. Questo permette di avere una visione completa delle prestazioni delle applicazioni e di intervenire rapidamente in caso di problemi. AWS CloudTrail, invece, fornisce un audit trail delle attività dell'account, contribuendo alla conformità e alla sicurezza, tracciando tutte le azioni effettuate sulle risorse AWS.

Per facilitare il deployment e la gestione delle risorse, AWS offre strumenti come AWS CloudFormation e AWS CodePipeline. AWS CloudFormation permette di modellare e impostare le risorse utilizzando file template, automatizzando il provisioning in modo programmabile e ripetibile. Ciò riduce il rischio di errori e consente di replicare facilmente ambienti di sviluppo, test e produzione. AWS CodePipeline, invece, è un servizio di integrazione e consegna continua che automatizza le fasi del processo di rilascio del software, facilitando lo sviluppo e il deployment di applicazioni.

2.2 Cloud Computing

Il Cloud Computing (Figura 2.2) consiste in un insieme di servizi fruibili dall'utente tramite una rete Internet a partire da un insieme di risorse preesistenti, configurabili e disponibili in remoto sotto forma di architettura distribuita. Tra questi servizi rientrano l'archiviazione, l'elaborazione e la trasmissione dei dati.

Per capire meglio la struttura e il funzionamento di questo strumento, analizziamo in dettaglio la sua architettura e i suoi componenti.

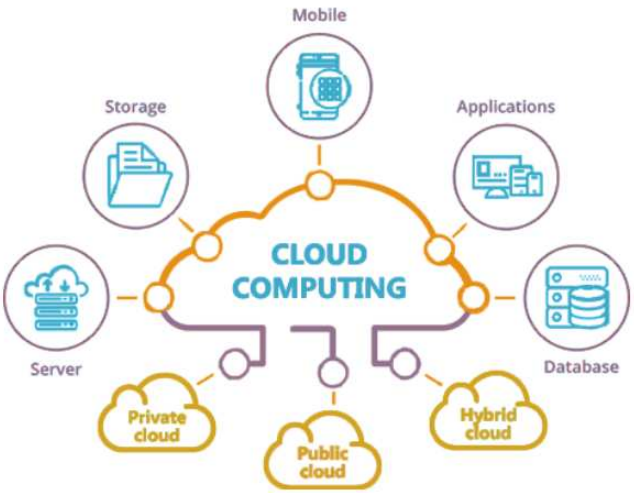


Figura 2.2: Cloud computing

2.2.1 Architettura

L'architettura del Cloud Computing prevede quattro componenti principali:

- *front-end*;

- *back-end*;
- *cloud*;
- *rete*.

Le piattaforme *front-end* sono quelle che interagiscono direttamente con l'utente finale, ossia le interfacce utente, le applicazioni lato client e il dispositivo o il Browser Web che facilita l'interazione degli utenti con i servizi di cloud computing e ne consente l'accesso.

Il *back-end*, invece, è la parte di architettura che gestisce la logica applicativa, l'elaborazione e l'archiviazione dei dati. Ne fanno parte il server, i database, gli storage e le applicazioni o i servizi che rispondono alle richieste del front-end.

La *rete* è, semplicemente, l'infrastruttura che permette la comunicazione tra front-end, back-end e il cloud. Può essere una *rete Internet*, *Intranet* o *Intercloud* che, rispettivamente, fanno riferimento ad una rete globale, ad una rete privata interna di un'organizzazione e all'interconnesione tra diversi cloud pubblici e privati.

Infine, il *cloud* è l'infrastruttura remota che ospita e gestisce i componenti di back-end e le risorse di elaborazione.

Con l'introduzione dei servizi cloud, il concetto di sicurezza è diventato fondamentale per proteggere dati, applicazioni e piattaforme. Assicurare la sicurezza del cloud richiede procedure e competenze diverse rispetto agli ambienti IT legacy.

Alcune delle misure di sicurezza da adottare sono:

- *crittografia dei dati*;
- *collaboratività*;
- *gestione dell'identità e degli accessi (IAM)*;
- *monitoraggio della sicurezza e della conformità*;
- *firewall e sicurezza della rete*.

Esistono tre tipologie fondamentali di modelli di servizi cloud (Figura 2.3), essi sono:

- *Software as a service (SaaS)*;
- *Platform as a service (PaaS)*;
- *Infrastructure as a service (IaaS)*.

2.2.2 Software as a Service

In una *struttura SaaS* un produttore mette a disposizione un programma mediante modalità telematiche, ovvero tramite il cloud. Essa consiste principalmente nell'utilizzo di programmi installati attraverso un server web.

Da specificare, però, che il modello SaaS non richiede necessariamente l'utilizzo di un'infrastruttura cloud, essendo nato prima della loro diffusione, anche se le sue prestazioni sono migliori con l'esclusivo utilizzo in cloud.

I vantaggi della piattaforma SaaS sono:

- *Riduzione dei costi iniziali*: la piattaforma SaaS riduce i costi di installazione ed implementazione delle infrastrutture ed elimina la necessità di middleware e hardware aggiuntivi;

- *Flessibilità e scalabilità*: essa permette alle aziende di rispondere rapidamente alle nuove opportunità e alle esigenze del mercato, facilitando ad esse l'aumento o la diminuizione delle risorse in base alle esigenze, senza dover investire in infrastrutture aggiuntive.
- *Implementazione rapida*: essa consente alle aziende di creare ed utilizzare nuovi software in tempi molto più brevi rispetto ai modelli tradizionali di distribuzione del software.
- *Accessibilità*: gli utenti possono accedere ai servizi SaaS in qualsiasi luogo e con qualsiasi dispositivo, tramite Internet, in modo da facilitare il lavoro remoto e la collaborazione tra team distribuiti geograficamente.

2.2.3 Platform as a Service

Il *modello PaaS* è stato pensato principalmente per sviuppatori e programmatori.

Infatti, questo modello fornisce piattaforme preconfigurate su cui l'utente può sviluppare, testare ed erogare applicazioni personalizzate.

Le piattaforme PaaS includono la gestione dei database, i sistemi di sicurezza, gli ambienti di sviluppo software, i diversi sistemi operativi e i server applicativi.

Amazon Web Services rientra in questa categoria.

I vantaggi della piattaforma PaaS sono i seguenti:

- *Riduzione dei costi*: i costi dei servizi PaaS vengono addebitati in base al consumo, così da evitare spese significative quando l'infrastruttura non è in uso.
- *Riduzione dei cicli di sviluppo delle applicazioni*: essi consentono di velocizzare lo sviluppo applicativo e di ridurre i tempi di distribuzione dei software.
- *Procedure DevOps efficienti*: essi facilitano lo sviluppo e il deployment delle app tramite la distribuzione continua.
- *Gestione delle misure di sicurezza*: le piattaforme PaaS offrono funzionalità di sicurezza integrate, come l'autenticazione, la gestione delle identità, la crittografia dei dati e il controllo degli accessi.
- *Produttività aumentata*: i team aziendali possono concentrarsi sulle iniziative strategiche, anziché sulle attività di gestione dell'infrastruttura, grazie all'accesso rapido agli strumenti e alle risorse necessarie.
- *Utilizzo delle competenze e degli investimenti esistenti*: gli sviluppatori hanno accesso a sistemi operativi, middleware, framework e altri strumenti di sviluppo usando i linguaggi di programmazione a loro familiari per scrivere codice in tempi brevi.

2.2.4 Infrastructure as a Service

L'*Infrastructure as a Service* è un servizio cloud in cui l'infrastruttura *IT* viene fornita agli utenti finali attraverso la rete Internet, eliminando la necessità di configurare i server di gestione. Il servizio IaaS è comunemente associato al *serverless computing*, ovvero tutte le attività di gestione dell'infrastruttura di back-end, tra cui provisioning, scaling, pianificazione e patch sul provider di cloud.

Questa architettura è supportata da *virtualizzazione*, *automazione* e *containerizzazione*.

Le *macchine virtuali (Virtual Machine, VM)* offrono ambienti completi su hardware fisico situato nei data center. Un hypervisor gestisce queste *VM*, separando le risorse hardware e permettendo il loro utilizzo da parte delle *VM* stesse.

L'*automazione IT* gestisce il deployment e la scalabilità delle risorse in modo ottimale, mentre, l'orchestrazione, ovvero l'esecuzione coordinata di più attività o processi di automazione dell'IT, automatizza diverse attività e configurazioni all'interno di gruppi di sistemi o macchine, assicurando un funzionamento efficiente.

Il *software*, con tutte le sue librerie, i framework e le dipendenze, viene "confezionato" in container, che sono più leggeri delle VM perché non includono un sistema operativo completo.

I vantaggi di IaaS sono i seguenti:

- *Economicità*: le risorse vengono utilizzate su richiesta e le aziende devono pagare solo per le risorse di calcolo, archiviazione e networking effettivamente utilizzate.
- *Efficienza*: le aziende che lo utilizzano riducono i ritardi legati all'espansione dell'infrastruttura e, in alternativa, non sperperano risorse per eccesso di capacità.
- *Produttività*: i reparti IT aziendali possono reindirizzare le risorse ad attività più strategiche, dato che il cloud provider è responsabile della configurazione e della manutenzione dell'infrastruttura fisica sottostante.
- *Affidabilità e scalabilità*: l'infrastruttura cloud offre ridondanza e tolleranza di errore integrate, con carichi di lavoro distribuiti su più server e strutture, quindi anche se un qualsiasi componente delle risorse hardware ha esito negativo, il servizio continuerà a funzionare.
- *Riduzione dei costi*: i costi IaaS sono abbastanza prevedibili e possono essere facilmente contenuti e preventivati.
- *Innovazione*: i team IT hanno più tempo da dedicare al lavoro strategico, ma l'IaaS consente di testare nuovi prodotti e idee in modo rapido ed economico.
- *Minore latenza*: la maggior parte dei provider di servizi cloud raggiunge una maggiore disponibilità e resilienza grazie all'aiuto di una rete globale che copre più aree geografiche. Posizionando app e servizi nelle regioni e nelle zone più vicine agli utenti finali, è possibile ridurre la latenza.

Nella figura 2.3 vengono riassunti i modelli di servizi di cloud computing.

2.2.5 Tipologie di Cloud Computing

Esistono quattro tipologie principali di cloud computing, ovvero *cloud pubblico*, *cloud privato*, *cloud ibrido* e *multicloud*.

Cloud pubblico

Il *cloud pubblico* è un ambiente cloud creato da un'infrastruttura IT non di proprietà dell'utente finale. Esso è un modello in cui le risorse di computing, come server e storage, sono di proprietà di un fornitore di servizi di cloud computing e sono condivise tra più clienti che possono accedere alle risorse via Internet.

Alcuni dei maggiori fornitori di cloud pubblico includono Amazon Web Services (AWS), Google Cloud, IBM Cloud e Microsoft Azure.

Le applicazioni cloud possono essere sviluppate utilizzando componenti infrastrutturali di base oppure impiegando servizi avanzati che offrono astrazioni delle operazioni di gestione, dell'architettura e della scalabilità dell'infrastruttura sottostante.

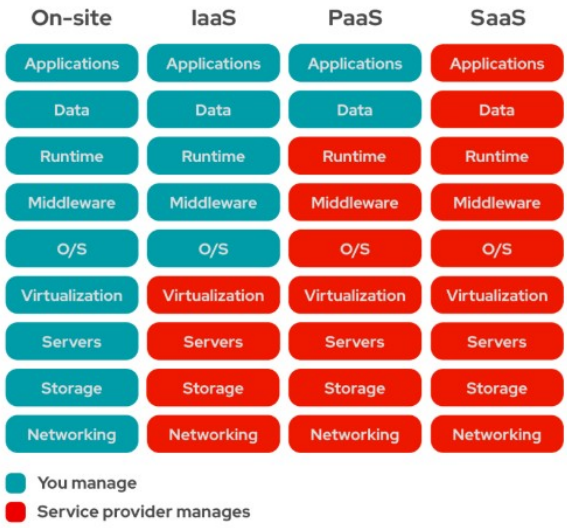


Figura 2.3: Modelli di servizi di Cloud Computing

Cloud privato

Il *cloud privato* consiste nell'implementazione delle risorse in locale, utilizzando strumenti di virtualizzazione e gestione delle risorse. È generalmente indicato per un singolo utente finale, in cui l'ambiente, di solito, viene eseguito all'interno del firewall di quell'utente o gruppo, permettendo un accesso esclusivo.

Cloud ibrido

Il *cloud ibrido* è un modello di cloud computing che combina elementi di cloud pubblico e cloud privato, permettendo alle organizzazioni di sfruttare i vantaggi di entrambi. In particolare, un cloud ibrido integra risorse di cloud pubblico con infrastrutture di cloud privato; ad esempio, i dati sensibili possono essere conservati in un cloud privato, mentre le risorse computazionali scalabili possono essere utilizzate nel cloud pubblico.

Multicloud

Il *multicloud* è un approccio al cloud computing che utilizza più servizi cloud da diversi fornitori. Questo, però, richiede una pianificazione attenta per gestire la complessità del servizio e garantire una gestione efficace.

2.3 Vantaggi di Amazon Web Services

Amazon Web Services è la piattaforma cloud più completa e utilizzata al mondo e offre molti vantaggi. Essi sono analizzati qui di seguito.

2.3.1 Scalabilità: Auto Scaling ed Elastic Load Balancing

La scalabilità si riferisce alla capacità della soluzione di connettività di crescere ed evolversi nel tempo al variare delle esigenze.

Amazon Web Services, grazie all'utilizzo di strumenti come *Auto Scaling* ed *Elastic Load Balancing*, ha permesso alle applicazioni di adattarsi dinamicamente alle esigenze variabili, adeguando facilmente risorse di calcolo e storage per garantire operatività continua.

Auto Scaling

L'Auto Scaling di AWS consente di adattare automaticamente la capacità delle risorse cloud in base alla domanda corrente, garantendo prestazioni ottimali e disponibilità continua delle applicazioni senza necessità di provisioning manuale.

AWS Auto Scaling (Figura 2.4) monitora le applicazioni e regola automaticamente la capacità per mantenere prestazioni stabili e prevedibili, riducendo al minimo i costi. Questo servizio permette di scalare diverse risorse su più servizi in pochi minuti tramite un'interfaccia utente semplice e potente, facilitando la creazione di piani di scalabilità per vari tipi di risorsa.

Per utilizzare l'Auto Scaling su AWS, è necessario creare un *endpoint SageMaker* come modello e definire una politica di scalabilità che aggiunge o rimuove istanze in base ai carichi di lavoro effettivi.

Le opzioni per l'Auto Scaling includono il monitoraggio degli obiettivi e le politiche di scalabilità dei passaggi. È fondamentale specificare i limiti di scalabilità, con un valore minimo di almeno 1 e un valore massimo che può essere illimitato.

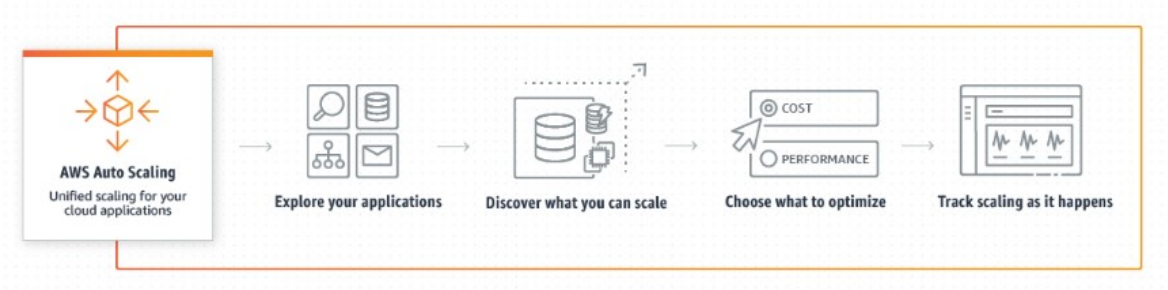


Figura 2.4: Funzionamento dell'Auto Scaling

Elastic Load Balancing (ELB)

Elastic Load Balancing di AWS è un servizio che distribuisce automaticamente il traffico in ingresso tra più destinazioni, come istanze *Amazon EC2*, container, indirizzi IP e funzioni *Lambda*.

ELB garantisce una distribuzione del carico efficiente ed affidabile, migliorando la disponibilità e la tolleranza ai guasti delle applicazioni. Inoltre, esso integra funzionalità di monitoraggio, sicurezza e scalabilità.

In riferimento alla sicurezza, esso si occupa di protezione delle applicazioni tramite la terminazione *SSL/TLS*, la gestione integrata dei certificati e l'autenticazione dei certificati client.

Tipi di ELB

Esistono le seguenti tipologie di ELB:

- *Application Load Balancer (ALB)* (Figura 2.5): esso è ottimizzato per bilanciare il carico delle applicazioni *HTTP* e *HTTPS* ed offre il routing operando sulle richieste, permettendo il routing avanzato basato su *URL*, host ed intestazioni *X-Forwarded-For*. Supporta *WebSockets* e *HTTP/2*.

- *Network Load Balancer (NLB)* (Figura 2.6): esso è progettato per gestire milioni di richieste al secondo, mantenendo bassa latenza. *NLB* opera a livello di trasporto, utilizzando indirizzi IP statici e routing basato sui protocolli *TCP/UDP*. È ideale per le applicazioni che richiedono prestazioni elevate e bassa latenza.

Figura 2.5: Application Load Balancer (ALB)

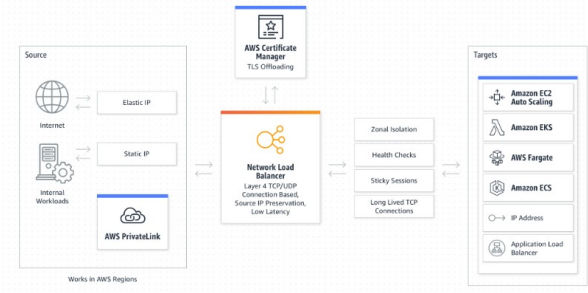


Figura 2.6: Network Load Balancer (NLB)

- *Gateway Load Balancer (GWLB)* (Figura 2.7): esso consente il bilanciamento del carico per le appliance di sicurezza virtuali, come firewall e sistemi di rilevamento delle intrusioni. Inoltre, offre il routing a livello di rete, facilitando l'inserimento di appliance di sicurezza nella tua architettura.

Figura 2.7: Gateway Load Balancer (GWLB)

2.3.2 Sicurezza

AWS supporta 143 standard di sicurezza e certificazioni di conformità, tra cui *PCI-DSS, HIPAA/HITECH, FedRAMP, GDPR, FIPS 140-2 e NIST 800-171*, aiutando i clienti a soddisfare i requisiti di conformità in tutto il mondo. AWS inoltre, offre un servizio chiamato *AWS Identity and Access Management* che gestisce le identità e l'accesso ai servizi e alle sue risorse. Questo servizio può essere usato anche da molti utenti contemporaneamente, in quanto è possibile

gestire gli utilizzatori connettendo le identità centralmente ad account AWS multipli. Esso usa credenziali di sicurezza e set di autorizzazioni temporanei per accedere alle risorse AWS.

In breve, grazie ad AWS Identity and Access Management (IAM), è possibile specificare chi o cosa può accedere alle risorse e ai servizi in AWS, gestire centralmente le autorizzazioni in modo molto dettagliato e analizzare gli accessi per affinare le autorizzazioni in AWS (Figura 2.8).

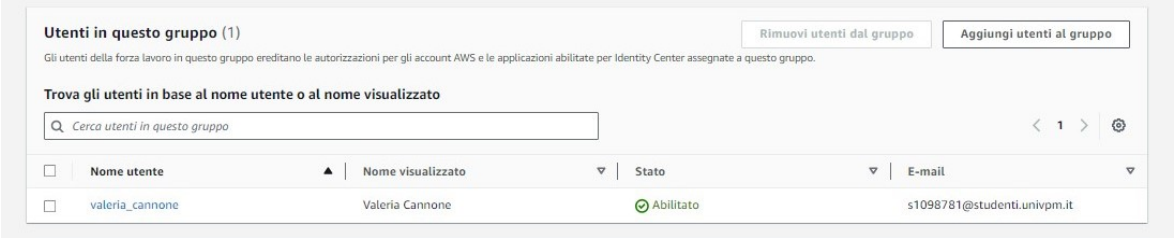


Figura 2.8: Gestione del gruppo associato ad utente IAM

Un altro strumento usato per la sicurezza, presente in AWS, è il servizio di gestione delle chiavi, chiamato AWS Key Management Service (AWS KMS) (Figura 2.9) che consente di creare, gestire e controllare le chiavi di crittografia tra le applicazioni e i servizi AWS.

L'accesso è regolato dalle policy delle chiavi, che permettono di controllare con precisione quando, chi e come può leggere i dati. Inoltre, utilizzando coppie di chiavi asimmetriche, è possibile creare e verificare firme digitali per garantire l'integrità dei dati. Le chiavi sono generate e protette in modo sicuro mediante dispositivi con moduli di sicurezza hardware, convalidati secondo lo standard statunitense *FIPS140* e considerati affidabili da istituzioni governative, finanziarie e sanitarie.

Un ulteriore vantaggio si ottiene integrando ad AWS KWS il servizio AWS Cloud Trail, il quale consente di registrare tutte le richieste API, comprese le azioni di gestione delle chiavi, gli eventi del ciclo di vita delle chiavi e l'utilizzo. Ciò aiuta a gestire i rischi, a soddisfare i requisiti di conformità o a condurre analisi forensi.

Oltretutto, AWS KMS è integrato con più di 100 servizi di AWS, tra cui tutti i principali database, gli strumenti per le analisi, l'archiviazione, la produttività e il flusso di lavoro.

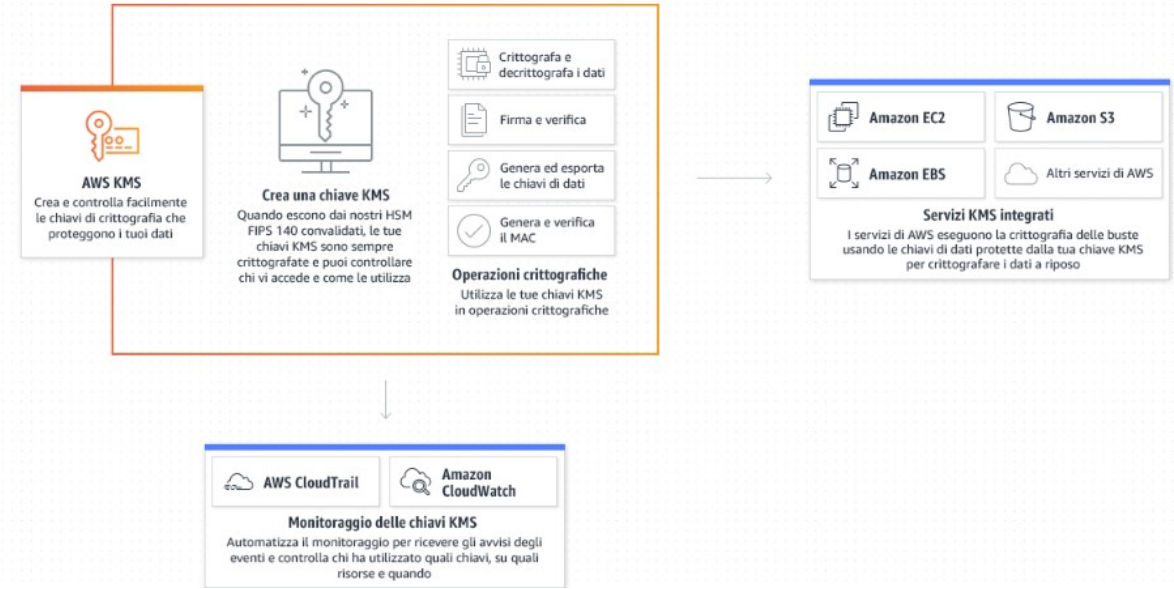


Figura 2.9: Funzionamento di AWS Key Management Service

2.3.3 Flessibilità

La piattaforma AWS offre anche flessibilità d'uso, permettendo di selezionare il sistema operativo, il linguaggio di programmazione, la piattaforma di applicazione web, database e altri strumenti essenziali tramite la Console di gestione AWS o API web service documentate.

2.3.4 Economicità

La piattaforma cloud si basa sul modello di prezzo *Pay-as-you-Go*, ovvero i costi dipendono esclusivamente dalle risorse utilizzate, calcolati in base all'effettivo utilizzo di storage e potenza di elaborazione.

AWS offre anche programmi di risparmio integrando sulla piattaforma cloud l'AWS Free Tier, ovvero il piano gratuito di Amazon Web Service. Le offerte variano in base alle tipologie di servizi, al numero di ore di utilizzo oppure in base alla quantità di memoria che è possibile usufruire su un determinato servizio; quest'ultimo fa riferimento ai servizi di storage, come, ad esempio, Amazon Bucket S3, il database online di AWS.

2.3.5 Affidabilità

Amazon Web Service opera in numerose regioni geografiche, ciascuna composta da più zone di disponibilità, ovvero cluster di data center distinti che offrono ridondanza e tolleranza ai guasti. Ogni data center AWS è progettato per garantire alta disponibilità e sicurezza, utilizzando hardware ridondanti e reti di alimentazione indipendenti.

2.3.6 Innovazione

Amazon AWS è innovativa perchè rilascia costantemente nuove funzionalità e servizi, rendendo essa stessa una piattaforma sempre più all'avanguardia rispetto ai suoi concorrenti. La piattaforma offre servizi all'avanguardia che operano in aree differenti, come Machine Learning, Intelligenza Artificiale, IoT, Blockchain ed Edge Computing.

In particolare, per il Machine Learning, Amazon Web Service offre il servizio *SageMaker* per la costruzione, l'addestramento e la distribuzione di modelli, mentre per accelerare lo sviluppo di modelli di Deep Learning è presente il servizio *AWS Deep Learning AMIs*. Altri servizi, come *AWS IoT Core* permettono la connessione sicura di dispositivi a Internet e l'interazione con altre applicazioni AWS; infine, *Amazon Managed Blockchain* facilita la creazione e la gestione di reti blockchain scalabili usando i framework *Hyperledger Fabric* ed *Ethereum*.

AWS è ritenuto innovativo soprattutto perché offre programmi di supporto per startup, tramite *AWS Activate*, fornendo risorse, training e crediti AWS per aiutare le nuove aziende a crescere. In aggiunta, mette a disposizione sull'AWS Marketplace, software, strumenti e soluzioni di terze parti che possono essere facilmente integrati con i servizi AWS, permettendo ai clienti di accedere a soluzioni innovative sviluppate dai partner, accelerando il time-to-market e la realizzazione dei progetti.

2.4 Principali servizi di AWS

I principali servizi offerti da AWS sono progettati per ottimizzare le operazioni aziendali e facilitare la trasformazione digitale. Essi si dividono in servizi per il computing, lo storage, i database, il networking, le analytics e il Machine Learing. Nel seguito analizziamo i principali di questi servizi:

2.4.1 Servizi per il computing

Tra i servizi per il computing possiamo citare *AWS EC2* e *AWS Lambda*.

AWS EC2

Amazon Elastic Compute Cloud (Amazon EC2) è un servizio che offre capacità di calcolo, sicuro e ridimensionabile per qualsiasi carico di lavoro nel cloud. Esso fornisce, in particolare, strategie di calcolo, memoria e archiviazione ottimizzate e tipi di istanze di calcolo ad alte prestazioni per fornire il miglior equilibrio, in ciascun settore, per differenti carichi di lavoro.

Amazon EC2 è supportato dai processori Intel, AMD, NVIDIA e AWS fornendo ulteriori ottimizzazioni anche per i costi.

Nello specifico, le istanze di uso generico sono ideali per le applicazioni che utilizzano le risorse a disposizione in pari proporzioni, come i server web e i repository di codice.

Le istanze per il calcolo ottimizzato sono ideali per le applicazioni che richiedono una notevole potenza di elaborazione e che possono beneficiare di processori ad alte prestazioni. A questa categoria appartengono i carichi di lavoro di elaborazione in batch, transcodifica dei supporti multimediali, i server web ad elevate prestazioni, il calcolo ad alte prestazioni (HPC), la modellazione scientifica, i server di videogiochi dedicati e motori di server pubblicitari, le applicazioni di Machine Learning e ad altre applicazioni a elevato utilizzo di calcolo.

Le istanze ottimizzate per la memoria sono progettate, invece, per erogare prestazioni ottimali riguardanti carichi di lavoro che elaborano grandi set di dati in memoria, come, ad esempio, database open source, cache in memoria e analisi dei Big Data in tempo reale.

Le istanze di calcolo accelerato su Amazon EC2 sono progettate per carichi di lavoro che beneficiano dell'accelerazione hardware, come GPU (Graphics Processing Unit) o FPGA (Field-Programmable Gate Array). Tali istanze possono offrire prestazioni significativamente superiori rispetto a quelle basate solo su CPU per specifiche applicazioni e carichi di lavoro. Esse sono utilizzate maggiormente in applicazioni di IA generativa, tra cui risposta a domande, generazione di codice, generazione di video e immagini e riconoscimento vocale, oppure in applicazioni HPC (High Performance Computing) nella ricerca farmaceutica, nell'analisi sismica, nelle previsioni meteorologiche e nella modellazione finanziaria.

Infine, le istanze di archiviazione ottimizzata sono progettate per offrire prestazioni elevate e un'elevata capacità di I/O (input/output) per applicazioni che richiedono un accesso rapido e scalabile ai dati. Tali istanze sono ideali per carichi di lavoro che necessitano di un'elevata velocità di lettura e scrittura dei dati, come database ad alte prestazioni e applicazioni con elevate esigenze di archiviazione. Alcuni di questi sono i database Amazon DynamoDB e MySQL, PostgreSQL, il servizio OpenSearch di Amazon e strumenti di analisi in tempo reale, come Apache Spark.

AWS Lambda

AWS Lambda è un servizio serverless di Amazon Web Services (AWS) che consente di eseguire codice senza dover gestire i server.

Con *Lambda* è possibile scrivere il codice che viene eseguito automaticamente in risposta ad eventi specifici, mentre AWS gestisce l'infrastruttura necessaria per l'esecuzione e la scalabilità del codice.

Esso è utile per estendere altri servizi AWS con logica personalizzata e per creare servizi di back-end scalabili e sicuri. Ad esempio, è possibile utilizzare il servizio per gestire l'inserimento di un elemento in un carrello su un sito web di e-commerce, eseguendo un codice personalizzato che elabora o aggiorna i dati in risposta a tale azione.

AWS Lambda, in particolar modo, può eseguire automaticamente il codice in risposta a vari eventi, come richieste HTTP tramite Amazon API Gateway e aggiornamenti di tabelle in Amazon DynamoDB o bucket Amazon S3, offrendo così un'ampia flessibilità per l'integrazione e l'automazione delle applicazioni.

2.4.2 Servizi per lo storage

Tra i servizi per lo storage citati in precedenza abbiamo il servizio Amazon S3. In aggiunta, saranno analizzati anche i servizi Amazon EBS e Amazon Glacier.

Amazon S3

Amazon S3 (Simple Storage Service) è un servizio di storage progettato per memorizzare e recuperare grandi volumi di dati, in qualsiasi momento e da qualsiasi luogo su Internet.

I clienti di tutte le entità e tutti i settori possono archiviare e proteggere una vasta gamma di dati per molteplici casi d'uso, come data lake, applicazioni native per il cloud e app mobili.

È possibile accedere ai dati da qualsiasi parte del mondo tramite una semplice richiesta HTTP/HTTPS con estrema sicurezza, poiché Amazon S3 utilizza policy di bucket, liste di controllo degli accessi (ACL) e autenticazione tramite AWS Identity and Access Management (IAM).

Inoltre, Amazon S3 offre opzioni di crittografia per proteggere i dati sia durante l'archiviazione sia durante il trasferimento e possiede funzioni di versioning e replicazione per garantire maggiore resilienza e disponibilità.

La flessibilità di *S3* lo rende adatto ad una vasta gamma di scenari, da applicazioni aziendali a soluzioni personali, offrendo una gestione centralizzata e la capacità di adattarsi automaticamente alle esigenze di storage.

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) è un servizio di archiviazione a blocchi scalabile, ad alte prestazioni e facile da utilizzare, progettato per essere usato con le istanze di Amazon EC2. Ogni volume EBS è progettato per funzionare come un'unità disco, permettendo letture e scritture rapide e affidabili. I volumi EBS possono essere utilizzati per diverse applicazioni, come database e file system, e possono essere ridimensionati facilmente in base alle esigenze. Questo servizio supporta anche snapshot per il backup, ovvero copie puntuali dei dati di un volume, la ripristinabilità dei dati; infine, esso offre opzioni di prestazioni variabili, tra cui SSD e HDD, per soddisfare diversi requisiti di carico di lavoro.

Amazon Glacier

Amazon S3 Glacier è un servizio di archiviazione a lungo termine fornito da AWS, progettato per conservare dati che non necessitano di accesso frequente, ma che devono essere mantenuti in modo sicuro e duraturo.

Il servizio è particolarmente vantaggioso grazie ai suoi costi molto bassi per l'archiviazione, rendendolo un'opzione economica rispetto ad altre soluzioni di storage. Ciò lo rende perfetto per dati di backup, archiviazione di dati storici e per la conformità normativa.

Infine, *S3 Glacier* offre diverse opzioni per il recupero dei dati, a seconda delle esigenze temporali. La modalità *Expedited* consente di recuperare i dati in pochi minuti, mentre la modalità *Standard* prevede un recupero in poche ore. Per grandi volumi di dati, la modalità di recupero più rapida è la *Bulk*, consentendo il recupero in un intervallo di 5-12 ore.

Inoltre, Glacier è progettato per integrarsi facilmente con Amazon S3 permettendo lo spostamento dei dati tra i due servizi, utilizzando le policy di *S3*.

2.4.3 Servizi per il database

I servizi per il database svolgono un ruolo cruciale nella gestione e nell'elaborazione dei dati. Amazon DynamoDB e Amazon RDS sono due servizi chiave che offrono soluzioni di database con caratteristiche distinte, adatte a diverse esigenze aziendali.

Amazon RDS

Amazon Relational Database Service (RDS) è un servizio che gestisce database relazionali e offre supporto per diversi motori di database come MySQL, PostgreSQL, MariaDB, Oracle e SQL Server.

Il servizio *RDS* semplifica la configurazione, la gestione e la scalabilità dei database relazionali, facilitando l'amministrazione, come il provisioning, la configurazione, la manutenzione e i backup, consentendo agli utenti di concentrarsi sullo sviluppo delle loro applicazioni. In particolar modo, è possibile concentrarsi sullo sviluppo delle applicazioni mentre la piattaforma Amazon Web Service gestisce tutta l'architettura del database.

Amazon DynamoDB

Amazon DynamoDB, invece, è un database NoSQL che si adatta automaticamente ai cambiamenti nel carico di lavoro. Questo facilita la gestione del volume di dati e delle richieste degli utenti senza richiedere configurazioni manuali.

Quando il traffico aumenta, *DynamoDB* amplia le risorse per mantenere prestazioni elevate; mentre, in caso contrario, riduce automaticamente la capacità del database per ottimizzare i costi. Questa caratteristica consente di mantenere l'efficienza operativa e di garantire che le applicazioni funzionino senza interruzioni o cali di prestazioni, indipendentemente dalle variazioni nella domanda.

2.4.4 Servizi per il networking

I servizi di networking di AWS forniscono una base solida per costruire architetture scalabili e sicure, supportando una vasta gamma di esigenze, dalla semplice configurazione di reti virtuali isolate alla gestione complessa di connettività tra più VPC e reti on-premises. Il servizio principale e più completo della piattaforma AWS è Amazon VPC.

Amazon VPC

Amazon Virtual Private Cloud (VPC) è un servizio essenziale di Amazon Web Services che consente di creare una rete privata virtuale, isolata all'interno del cloud AWS. Questo servizio offre la possibilità di configurare e gestire un ambiente di rete sicuro e altamente personalizzabile, simile a una rete tradizionale on-premises, ma con tutti i vantaggi della scalabilità e della flessibilità del cloud.

Con Amazon VPC è possibile suddividere la rete in *subnet*, che possono essere configurate come pubbliche o private. Le subnet pubbliche sono accessibili da Internet, mentre quelle private rimangono isolate. Ogni VPC è assegnata ad un intervallo di indirizzi IP privati e le tabelle di routing consentono di gestire il traffico tra le subnet e l'esterno, incluso Internet e altre reti, attraverso i gateway.

Per le risorse nelle subnet private, VPC offre l'uso di *NAT Gateway* o *NAT Instance* per permettere l'accesso a Internet senza esporre direttamente le risorse.

La sicurezza è garantita tramite *Security Groups* e *Network ACLs (Access Control Lists)*, che offrono un controllo dettagliato del traffico.

Amazon VPC consente anche la connessione tra VPC tramite *peering*, facilitando la comunicazione tra reti separate. Per collegare una VPC ad una rete on-premises si può utilizzare una *Virtual Private Network (VPN)*, creando un canale sicuro e criptato attraverso Internet. Gli endpoint VPC migliorano ulteriormente la sicurezza e le prestazioni, consentendo alle risorse di accedere ai servizi AWS, come Amazon S3 e DynamoDB, senza l'uso di Internet.

Il *Transit Gateway* semplifica la gestione della connettività tra più VPC e reti on-premises fornendo un punto di collegamento centralizzato. Questo servizio facilita la connessione e l'interoperabilità tra diverse reti, ottimizzando la rete nel suo complesso.

In sintesi, Amazon VPC offre un ambiente di rete isolato e sicuro con un controllo completo e una flessibilità elevata, essenziale per costruire e gestire architetture di rete complesse nel cloud AWS.

Amazon CloudFront

Amazon CloudFront è un servizio di *Content Delivery Network (CDN)* altamente scalabile e sicuro che distribuisce contenuti a livello globale con bassa latenza e alta velocità di trasferimento.

Esso è progettato per integrarsi perfettamente con altri servizi AWS. Inoltre, CloudFront accelera la distribuzione di dati, video, applicazioni e API ai clienti in tutto il mondo, garantendo un'esperienza utente ottimale.

CloudFront utilizza una rete globale di *edge locations* per memorizzare nella cache le informazioni collocandole in data center più vicini agli utenti finali. Questo riduce la distanza che i dati devono percorrere, migliorando i tempi di risposta e la disponibilità del servizio.

Il servizio supporta soprattutto la distribuzione sia di contenuti statici che dinamici, rendendolo ideale per siti web, applicazioni mobili, streaming video e distribuzione di file di grandi dimensioni.

Esso dispone di metodi di sicurezza avanzata; infatti, include il protocollo HTTPS, la protezione DDoS integrata tramite *AWS Shield*, l'autenticazione tramite *AWS IAM* e l'integrazione con *AWS WAF* per la protezione contro attacchi web.

2.4.5 Servizi per le analytics

Nel campo dell'analisi dei dati, AWS mette a disposizione una serie di strumenti avanzati che aiutano le aziende ad ottenere informazioni preziose dai loro dati. Tra questi, Amazon Kinesis e Amazon Athena sono rilevanti per questo tipo di operazioni.

Amazon Kinesis

Amazon Kinesis è una console di servizi per la gestione e l'elaborazione di dati in tempo reale. I servizi Kinesis permettono di raccogliere, elaborare e analizzare flussi di dati in tempo reale, consentendo alle aziende di ottenere informazioni istantanee e di prendere decisioni rapide basate su dati aggiornati. Le principali tipologie di AWS Kinesis sono:

- *Amazon Kinesis Data Streams*: esso consente di raccogliere ed elaborare dati di streaming ad alta velocità. I dati sono suddivisi in *shard*, sequenze di record di dati, che possono essere letti e scritti in parallelo. Le applicazioni possono connettersi a questi stream per processare i dati in tempo reale, ad esempio per il monitoraggio, l'analisi e la gestione delle attività operative.

- *Amazon Kinesis Video Stream*: esso è ideale per casi d'uso come la videosorveglianza, la gestione delle telecamere e l'analisi di immagini in tempo reale. Supporta la registrazione e l'analisi di video, nonché l'integrazione con servizi di Machine Learning per estrarre informazioni dai flussi video.

Amazon Athena

Amazon Athena è un servizio di query interattive serverless basato su framework open source, che supporta formati di file e tabelle aperti. Ad esempio, consente di analizzare i dati direttamente in Amazon S3 utilizzando SQL standard, senza la necessità di gestire infrastrutture o configurare cluster. Questo lo rende particolarmente adatto per analisi specifiche e per query dirette sui dati già archiviati in S3.

Athena si adatta ad una vasta gamma di casi d'uso, tra cui l'analisi dei log, la reportistica aziendale e il monitoraggio delle attività. Su AWS, è possibile inoltrare query utilizzando la console del servizio, definendo uno schema attraverso la procedura guidata o immettendo istruzioni *DDL, Data Definition Language*. L'editor di query integrato consente di scansionare i dati e di compilare il catalogo con definizioni di tabella e partizioni nuove o modificate.

I risultati delle query saranno visualizzati nella console in pochi secondi e automaticamente salvati in un percorso personalizzato in S3 o scaricabili in locale. Questo sistema facilita l'analisi di grandi quantità di dati, consentendo di applicare le proprie conoscenze di SQL in modo efficace e immediato.

2.4.6 Servizi per il Machine Learning

In questo contesto, *Amazon SageMaker* e *Amazon Comprehend* sono due dei principali servizi di Machine Learning di AWS che offrono capacità distintive e complementari.

Amazon SageMaker

Amazon SageMaker consente agli sviluppatori e ai data scientist di costruire, addestrare e distribuire modelli di Machine Learning.

Esso fornisce strumenti e ambienti integrati; in particolare, include *Jupyter Notebooks* preconfigurati per la preparazione dei dati, l'esplorazione e la creazione dei modelli. Gli utenti possono avviare l'addestramento su istanze GPU o CPU, ottimizzando le risorse in base alle esigenze. Il servizio gestisce automaticamente la distribuzione delle risorse e la scalabilità durante l'addestramento, riducendo i tempi e i costi associati.

In particolare, il servizio supporta l'implementazione grazie ai 150 modelli open source popolari presenti sulla piattaforma, come l'elaborazione del linguaggio naturale, il rilevamento di oggetti e i modelli di classificazione delle immagini.

Una volta addestrato, il modello può essere facilmente distribuito in produzione utilizzando *SageMaker Endpoint*, che consente previsioni in tempo reale, oppure come *Batch Transform*, per fare previsioni su grandi volumi di dati.

Amazon Comprehend

Infine, *Amazon Comprehend* è un servizio di elaborazione del linguaggio naturale (NLP) che utilizza il Machine Learning per estrarre informazioni dai testi.

Comprehend è in grado di identificare entità, comprendere il sentiment, analizzare le frasi e, persino, estrarre le relazioni tra i concetti in un testo.

Questo servizio consente alle aziende di automatizzare l'analisi del contenuto e ottenere informazioni significative dai dati testuali. Inoltre, Comprehend supporta anche il rilevamento di tematiche principali e la traduzione automatica, ampliando le possibilità di analisi dei dati e migliorando la comprensione del linguaggio naturale.

CAPITOLO 3

Esperienze sul riconoscimento di immagini

In questo capitolo studieremo cos'è e come funziona il riconoscimento di immagini analizzando il servizio cloud di AWS, Amazon Rekognition e le sue caratteristiche più importanti.

Inoltre, dimostreremo il suo funzionamento in uno studio effettuato sulla piattaforma applicando, su un dataset di immagini, la tecnica Custom Labels per le etichette personalizzate. Infine, esamineremo i casi d'uso di Rekognition e come viene impiegato nei vari settori della vita reale.

3.1 Introduzione al riconoscimento di immagini

Il riconoscimento di immagini, o *image recognition*, è un processo che serve ad identificare un oggetto o una caratteristica in un'immagine o un video. In informatica, esso appartiene ad una branca dell'Intelligenza Artificiale detta *Computer Vision*, o *visione artificiale*.

L'image recognition velocizza compiti ripetitivi e analizza le immagini più rapidamente e con maggiore accuratezza rispetto all'ispezione manuale. Questa tecnica è essenziale per numerose applicazioni ed è l'elemento principale delle applicazioni di Deep Learning, tra cui:

- *Ispezione visiva*: essa consente di individuare rapidamente le parti difettose durante la produzione, esaminando migliaia di componenti su una linea di montaggio con grande velocità.
- *Classificazione di immagini*: questa funzionalità permette di categorizzare le immagini in base al loro contenuto, risultando particolarmente utile in applicazioni come il recupero di immagini e i sistemi di raccomandazione nell'e-commerce.
- *Guida autonoma*: essa è fondamentale per le applicazioni di guida autonoma, come la capacità di riconoscere segnali di stop o pedoni in un'immagine.
- *Robotica*: i robot possono utilizzare il riconoscimento delle immagini per identificare oggetti e migliorare la navigazione autonoma, riconoscendo luoghi e oggetti lungo il loro percorso.

3.1.1 Amazon Rekognition

Amazon Rekognition è un servizio di Computer Vision offerto da Amazon Web Services (AWS) presente sulla piattaforma dal 2016. Esso contiene algoritmi pre-addestrati per l'analisi dei media, le cosiddette *demo*.

Questo servizio è basato sul modello cloud SaaS ed utilizza tecniche avanzate di Deep Learning per analizzare immagini e video, rendendo possibile l'identificazione di oggetti, persone, testi, scene e attività, senza la necessità di conoscenze approfondite di Machine Learning. La sua scalabilità e la sua semplicità d'uso lo rendono, quindi, accessibile ad una vasta gamma di utenti.

Inoltre, Amazon Rekognition è uno strumento utilizzato da vari enti privati e da numerose agenzie governative degli Stati Uniti, in quanto ritenuto altamente valido anche per quanto riguarda le analisi sulla sicurezza.

Tuttavia, nonostante l'ampio utilizzo dei sistemi di riconoscimento facciale per diverse applicazioni, è fondamentale considerare le implicazioni etiche legate ai pregiudizi razziali e di genere, che possono influenzare l'affidabilità di questi sistemi, causando diverse controversie.

3.1.2 Controversie

Nel 2018, i ricercatori del Massachusetts Institute of Technology (MIT) Joy Buolamwini e Timnit Gebru hanno pubblicato uno studio intitolato *Gender Shades*. Questo studio consisteva nel raccogliere un set di immagini in cui i volti erano etichettati con informazioni sulla posizione del viso, sul genere e sul tono della pelle. Le immagini sono state analizzate tramite piattaforme di riconoscimento facciale SaaS di Face++, IBM e Microsoft.

I risultati hanno rivelato che tutti e tre i classificatori funzionavano meglio sui volti maschili, con tassi di errore sui volti femminili superiori, dall'8,1% al 20,6%. Inoltre, le prestazioni peggioravano ulteriormente sui volti femminili con tonalità di pelle più scura, con tassi di errore che variavano dal 20,8% al 30,4%.

Gli autori hanno ipotizzato che questa discrepanza fosse dovuta, principalmente, alla predominanza di immagini di maschi chiari nei dati di addestramento di Megvii, IBM e Microsoft, evidenziando una distorsione nei set di dati utilizzati.

Nel gennaio 2019, i ricercatori Inioluwa Deborah Raji e Joy Buolamwini hanno pubblicato un documento di follow-up, ripetendo l'esperimento sulle versioni aggiornate delle stesse tre piattaforme e aggiungendo due ulteriori sistemi: *Kairos* e *Amazon Rekognition*. Sebbene i tassi di errore complessivi fossero migliorati rispetto all'anno precedente, tutti e cinque i sistemi continuavano a mostrare prestazioni migliori sui volti maschili rispetto ai volti femminili con tonalità di pelle scura. Questo studio sottolinea la necessità di affrontare e mitigare le distorsioni presenti nei sistemi di riconoscimento facciale, per garantire l'equità e l'affidabilità delle tecnologie utilizzate.

3.1.3 Come funziona Amazon Rekognition

Amazon Rekognition fornisce, in particolare, due set di API per l'analisi visiva, ovvero:

- *Amazon Rekognition Image*, per l'analisi delle immagini;
- *Amazon Rekognition Video*, per l'analisi dei video.

Amazon Rekogniton Image si occupa del rilevamento di oggetti, scene e concetti nelle immagini. Esso riconosce le celebrità, rileva il testo in diverse lingue nonchè contenuti o immagini espliciti, inappropriati o violenti.

Inoltre, rileva, analizza e confronta volti e caratteristiche facciali, come età ed emozioni. Infine, rileva la presenza di DPI (dispositivi di protezione individuale).

Esso include, anche, il miglioramento delle app fotografiche, la catalogazione di immagini e la moderazione dei contenuti, fornendo, così, un utilizzo più ampio e vantaggioso per gli utenti.

Amazon Rekognition Video, invece, è in grado di tracciare persone e oggetti attraverso i fotogrammi video, riconoscere gli oggetti, riconoscere le celebrità e cercare video archiviati e in streaming per persone di interesse.

Inoltre, così come Amazon Rekognition Image, analizza i volti per esaminare gli attributi di un soggetto, ovvero età, emozioni e genere, rilevare contenuti o immagini espliciti, inappropriati o violenti per aumentare la sicurezza. Infine, esso aggrega ed ordina i risultati delle analisi in ordine temporale e per segmenti, e si occupa di rilevare persone, animali domestici e pacchi nei video in streaming.

3.1.4 Applicazioni pratiche di Amazon Rekognition

Amazon Rekognition trova applicazione in una vasta gamma di settori, migliorando l'efficienza e l'accuratezza di diversi processi. Alcuni esempi concreti di utilizzo includono:

- *Sicurezza e sorveglianza*: le forze dell'ordine e le agenzie di sicurezza utilizzano Amazon Rekognition per identificare sospetti e persone scomparse analizzando le immagini e i video delle telecamere di sorveglianza. Ciò consente di rispondere più rapidamente alle emergenze e migliorare la sicurezza pubblica.
- *E-commerce e pubblicità*: le piattaforme di e-commerce possono utilizzare il servizio AWS per migliorare le raccomandazioni di prodotti. Ad esempio, analizzando le immagini dei prodotti, il sistema può suggerire ai clienti articoli simili, aumentando, così, le vendite e la soddisfazione degli utenti.
- *Gestione dei media*: le aziende che gestiscono grandi volumi di contenuti multimediali, come foto e video, possono utilizzare Amazon Rekognition per catalogare ed organizzare i loro archivi. Ciò facilita la ricerca e il recupero di contenuti specifici, migliorando l'efficienza operativa.
- *Sanità*: gli ospedali e le cliniche possono impiegare Amazon Rekognition per analizzare le immagini mediche, come radiografie e risonanze magnetiche, per rilevare anomalie e supportare la diagnosi precoce di malattie.
- *Intrattenimento e media*: i produttori di contenuti video possono utilizzare Amazon Rekognition per analizzare filmati e creare metadati dettagliati. Ciò è utile per la creazione di riassunti automatici, per la ricerca di scene specifiche e per la gestione dei diritti digitali.

Nel seguito, esamineremo casi di studio specifici che illustrano l'uso di Amazon Rekognition in questi settori applicativi.

3.2 Funzionalità principali di Amazon Rekognition

Tra le capacità di visione artificiale pre-addestrate e personalizzabili, la piattaforma fornisce le seguenti funzionalità principali:

3.2.1 Moderazione dei contenuti

La moderazione dei contenuti rileva contenuti potenzialmente non sicuri, inappropriati o indesiderati in immagini e video.

Le API di Rekognition per questo determinato servizio sono utilizzate, soprattutto, nei social media, nei media radiotelevisivi, nella pubblicità e nelle situazioni di e-commerce per

creare un'esperienza utente più sicura, per fornire garanzie di sicurezza del marchio agli inserzionisti e per rispettare le normative locali e globali.

Nell'attuale panorama digitale, molte aziende si affidano ai moderatori umani per esaminare contenuti generati da terze parti o dagli utenti. Tuttavia, questo approccio presenta delle limitazioni significative. I moderatori umani, infatti, non possono adattarsi in modo efficiente alla crescente domanda di monitoraggio, il che comporta spesso esperienze utente di scarsa qualità, costi elevati o, nei casi peggiori, danni alla reputazione del marchio. Proprio per questo motivo, le API di Rekogntion risultano particolarmente utili, in quanto, grazie ad esse, i moderatori umani possono concentrarsi su un volume molto più ridotto di contenuti, generalmente tra l'1% e il 5% del totale, già identificato dall'IA come potenzialmente problematico. Questo approccio consente loro di dedicarsi ad attività di maggiore valore, garantendo, allo stesso tempo, una copertura di moderazione completa a costi significativamente ridotti.

Per ottenere questo risultato, il sistema definisce tre livelli di moderazione (L1, L2, L3) per etichettare le categorie di contenuti inappropriati, indesiderati o offensivi, facilitando, così, una classificazione gerarchica ed efficace del contenuto. In particolare:

- *L1 (Low)*: essa rappresenta la classificazione più bassa e generalmente include materiale che potrebbe non essere offensivo, ma potrebbe comunque essere considerato problematico da alcuni utenti. Questa categoria identifica contenuti che potrebbero includere comportamento irrispettoso o immagini che presentano contenuti ambigui e che richiedono ulteriori verifiche.

- *L2 (Moderate)*: questa categoria comprende contenuti che sono chiaramente più problematici rispetto ai contenuti di livello L1. Essi sono esplicitamente inappropriati e possono includere elementi che violano le linee guida comunitarie o che sono potenzialmente dannosi. In dettaglio, tale categoria contiene immagini raffiguranti comportamenti violenti o contenuti sessuali non espliciti, ma comunque riconducibili a violazioni delle linee guida.

- *L3 (High)*: il livello L3 include contenuti altamente problematici o esplicitamente inappropriati. Questo livello di severità indica contenuti che sono chiaramente offensivi o violano severamente le politiche di moderazione. Esso contiene contenuti estremamente violenti, contenuti sessuali espliciti o che incitano all'odio o alla discriminazione.

Nella Figura 3.1 è possibile osservare l'immagine di una famiglia in spiaggia. I risultati ottenuti, secondo la moderazione dei contenuti, segnalano la presenza di nudismo non esplicito di livello L2 ed elementi di livello L1 (costumi da bagno).

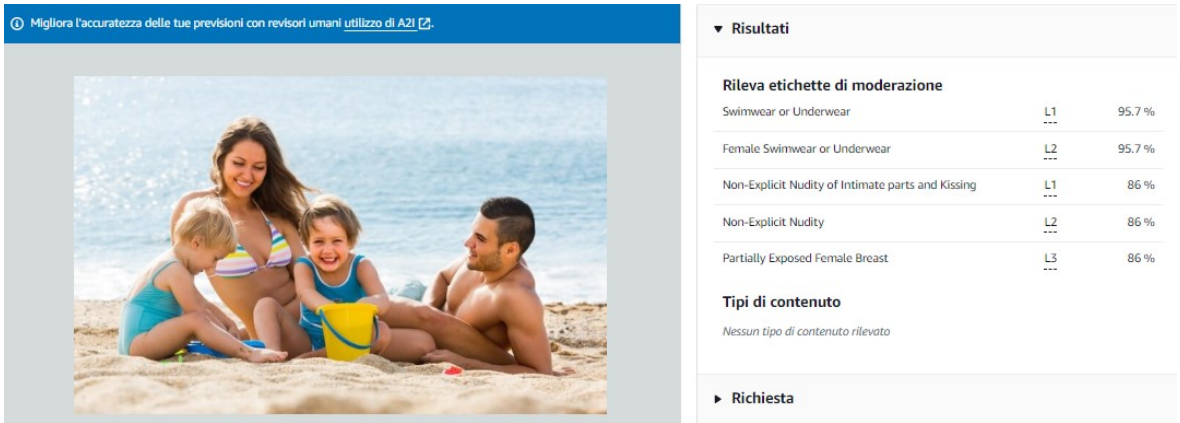


Figura 3.1: Moderazione dei contenuti

Inoltre, la funzione di filtraggio dei contenuti è in grado di identificare se il tipo di contenuto è animato, come cartoni animati, fumetti, manga e anime, oppure illustrato, come disegni, dipinti e schizzi.

3.2.2 Biometria facciale

Le funzioni sulla biometria facciale di Amazon Rekognition includono:

- rilevamento e analisi facciale;
- confronto e ricerca facciale;
- riconoscimento facciale.

Rilevamento e analisi facciale

Amazon Rekognition offre potenti strumenti per rilevare e analizzare volti nelle immagini e nei video. Una delle sue funzionalità più avanzate è la possibilità di memorizzare e cercare corrispondenze di volti in una raccolta.

Per iniziare, è necessario creare una raccolta di immagini di volti in diverse posizioni; a questo punto, esso è in grado di rilevare e riconoscere attributi, come occhi aperti, occhiali, barba o baffi, per ciascuno dei soggetti raffigurati.

Supponiamo di fornire un'immagine e di avviare l'analisi. Dai risultati (Figura 3.2), possiamo notare un grande quantitativo di attributi che il sistema è stato in grado di rilevare dal volto proposto, come, ad esempio, il sesso, l'età, le emozioni e le caratteristiche fisiche.

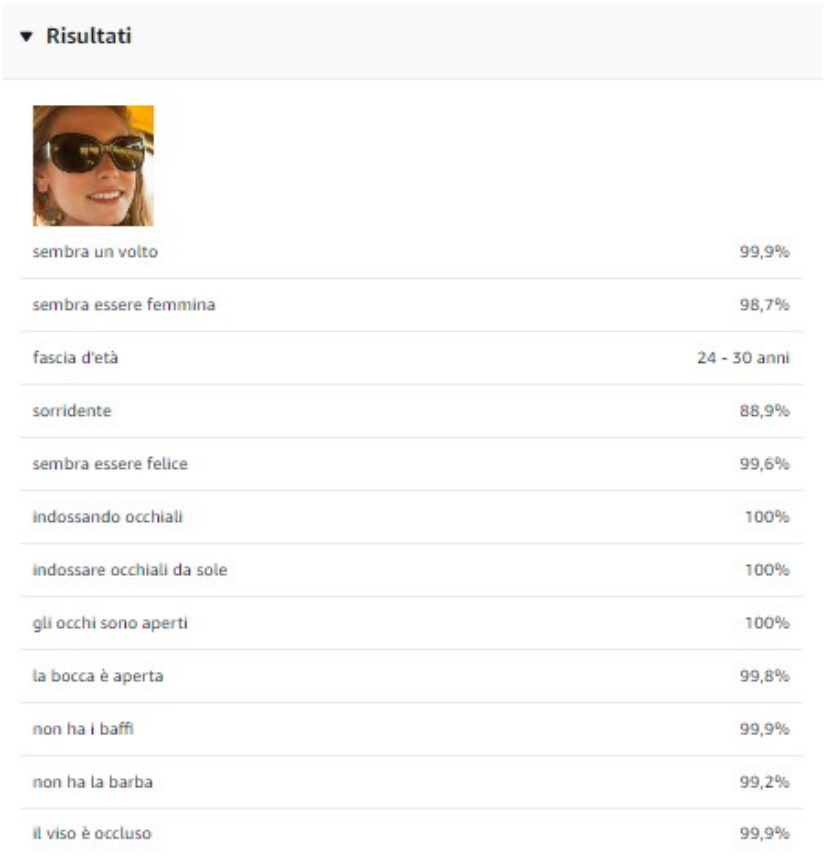


Figura 3.2: Risultati ottenuti dall'applicazione della demo sull'analisi facciale

Amazon Rekognition Face Liveness

La funzione *Amazon Rekognition Face Liveness*, o *riconoscimento facciale*, si occupa di analizzare il viso di utenti reali ed è utilizzato, maggioramente, per rilevare in pochi secondi eventuali malintenzionati che tentano di utilizzare contraffazioni durante la verifica facciale. Tra i principali casi d'uso sono inclusi:

- la verifica di utenti fraudolenti;
- un sistema di autenticazione avanzata;
- la verifica dell'età di un utente;
- il rilevamento dei bot.

Confronto e ricerca facciale

In aggiunta, Amazon Rekognition è in grado di identificare punti di riferimento facciali (ad esempio, posizione degli occhi), rilevare emozioni (come felicità o tristezza) ed altri attributi (ad esempio, presenza di occhiali, occlusione del viso). Nel momento in cui il sistema rileva un volto, esso analizza gli attributi facciali e restituisce un punteggio di affidabilità per ciascun attributo, chiamato *F1 Score*.

I modelli di rilevamento dei volti utilizzati da Amazon Rekognition Image e Amazon Rekognition Video non supportano il rilevamento dei volti in personaggi dei cartoni animati o entità non umane. Ciò è possibile attraverso la funzione Custom Labels, funzionalità che verrà spiegata di seguito.

In merito al confronto facciale, noto come *Face Comparison*, è possibile sottoporre al modello di apprendimento automatico due immagini contenenti uno o più volti. Il sistema li analizza e confronta, fornendo una valutazione dettagliata per ogni coppia di volti trovata nelle due foto. Inoltre, restituisce una percentuale di somiglianza che indica quanto il modello ritiene che le due immagini rappresentino la stessa persona.

Nella Figura 3.3 è rappresentato il risultato del confronto facciale ottenuto dalla demo di AWS.

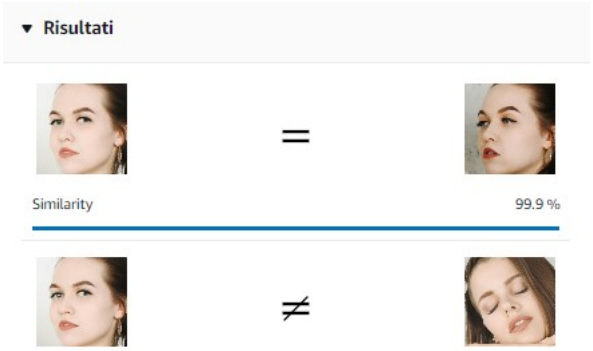


Figura 3.3: Cattura del risultato di una Facial Comparison effettuata su due immagini rappresentanti varie ragazze

Questa tecnologia è abbastanza avanzata da riconoscere la stessa persona anche se le immagini sono state scattate a distanza di molti anni. Ad esempio, se si forniscono due immagini della stessa persona scattate a distanza di anni, lo strumento può identificare correttamente che si tratta dello stesso individuo con una percentuale di somiglianza del 99,9% (Figura 3.4).

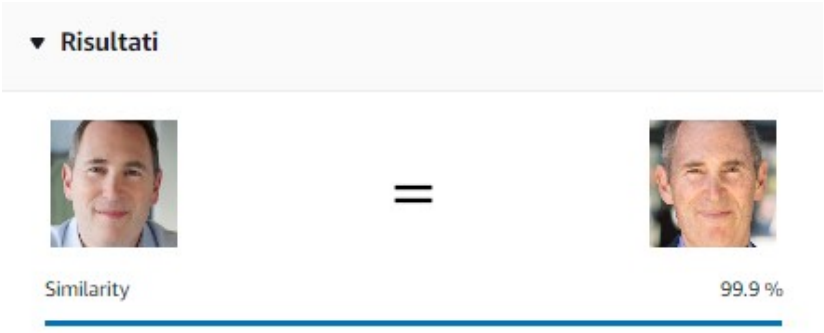


Figura 3.4: Confronto di due immagini della stessa persona a distanza di anni

3.2.3 Custom Labels

Amazon Rekognition Custom Labels consente di creare modelli di Machine Learning personalizzati per rilevare oggetti, scene o concetti specifici nelle immagini, che non sono coperti dalle etichette generiche predefinite, rispondendo così ad esigenze applicative particolari.

Grazie all'integrazione con altri servizi AWS, come Amazon S3 per l'archiviazione delle immagini e AWS Lambda per l'elaborazione serverless, è possibile gestire l'intero processo di creazione del modello senza competenze specifiche di apprendimento automatico, grazie alle funzionalità AutoML (Machine Learning automatico). Inoltre, è possibile etichettare manualmente i media o, per dataset di grandi dimensioni, utilizzare il servizio Amazon SageMaker Ground Truth.

Dopo l'addestramento, le prestazioni del modello possono essere valutate tramite metriche dettagliate, ovvero la Precision, il Recall, l'F1 Score e le previsioni comparative. Il modello può essere utilizzato immediatamente per l'analisi delle immagini, con la possibilità di iterare e riaddestrare lo stesso, così da migliorare le prestazioni. I feedback sui risultati delle previsioni aiutano a perfezionare il modello.

Amazon Rekognition offre due funzionalità principali:

- *Amazon Rekognition Image label detection*: esso è uno strumento ideale per identificare etichette comuni in immagini e video, senza la necessità di addestrare un modello;
- *Amazon Rekognition Custom Labels*: esso è consigliato per esigenze specifiche che richiedono il riconoscimento di etichette non comuni o personalizzate.

Esempi di utilizzo di Custom Labels includono l'identificazione di:

- loghi su immagini dei social media;
- parti di macchine su linee di assemblaggi;
- personaggi dei cartoni animati;
- prodotti di specifici marchi sugli scaffali dei negozi;
- prodotti agricoli, valutando e classificando la loro qualità.

Nella Figura 3.5 è illustrata un'immagine analizzata con la demo di Rekognition per il rilevamento delle etichette. Questa immagine mostra come la funzione Custom Labels ha identificato la presenza di persone e di un'opera architettonica, indicando che questi elementi si trovano all'aperto e riconoscendo la presenza di piante e prato. Inoltre, sullo sfondo, il sistema etichetta la presenza di edifici.

L'analisi ha restituito una percentuale di risultato compresa tra il 95% e il 100%, indicando un riconoscimento molto accurato.

Figura 3.5: Dimostrazione della funzione Custom Labels di Rekognition

3.2.4 Rilevamento dei testi

Il *rilevatore di testi* di Rekognition rileva e riconosce elementi testuali all'interno di immagini o video.

L'uso del rilevatore di testi nelle immagini può essere applicato in diverse soluzioni innovative:

- *Ricerca visiva*: questa tecnologia consente di individuare e mostrare immagini contenenti lo stesso testo, facilitando la ricerca e il recupero di immagini specifiche in grandi database visivi.
- *Analisi dei contenuti*: attraverso l'estrazione del testo dai fotogrammi video, è possibile ottenere approfondimenti su temi ricorrenti, come notizie, punteggi sportivi, numeri di giocatori o sottotitoli. Le applicazioni possono esaminare il testo riconosciuto per identificare contenuti rilevanti.
- *Navigazione assistita*: essa è utile per le persone con disabilità visive. Infatti, utilizzare il riconoscimento del testo è utile per identificare nomi di ristoranti, negozi o segnali stradali (come nella Figura 3.6). Ciò migliora l'accessibilità e l'orientamento in ambienti esterni.
- *Sicurezza pubblica e gestione del traffico*: il rilevamento automatico delle targhe dei veicoli tramite telecamere stradali può essere utilizzato per migliorare la sicurezza stradale e per scopi di controllo del traffico.
- *Protezione dei dati personali*: le soluzioni di filtraggio possono identificare e rimuovere informazioni personali presenti nelle immagini, contribuendo a proteggere la privacy e a rispettare le normative vigenti.

Per quanto riguarda il rilevamento del testo nei video, le possibili applicazioni includono:

- *ricerca di contenuti video*: esse possono essere progettate per trovare segmenti video specifici contenenti parole chiave, rendendo più semplice la ricerca all'interno di archivi video;

- *moderazione dei contenuti video*: è possibile rilevare testi inappropriati, linguaggio offensivo o spam nei video per assicurarsi che il contenuto rispetti le politiche aziendali e comunitarie;
- *gestione delle sovrimpressioni di testo*: la tecnologia può individuare e tracciare le sovrapposizioni di testo nei video, consentendo la sostituzione del testo per la localizzazione o il miglioramento della presentazione visiva;
- *allineamento grafico preciso*: identificando la posizione esatta del testo nei video, è possibile allineare con precisione altri elementi grafici, migliorando la qualità complessiva della produzione video.

Figura 3.6: Esempio di applicazione della funzione di rilevamento dei testi in un'immagine

3.2.5 Riconoscimento di volti celebri

Un'ulteriore funzionalità di Amazon Rekognition, oltre al riconoscimento facciale spiegato in precedenza, è il riconoscimento di volti celebri. Queste funzionalità presentano alcune differenze fondamentali nei loro casi d'uso e nelle best practice.

Il riconoscimento delle celebrità è pre-addestrato con la capacità di riconoscere centinaia di migliaia di persone famose in campi come sport, media, politica e affari.

Nella Figura 3.7, Amazon Rekognition è stato in grado di riconoscere che la persona raffigurata nell'immagine è Jeff Bezos, il fondatore di Amazon, con una percentuale di 99,7%. Inoltre, la console permette di visualizzare informazioni aggiuntive su di lui effettuando una ricerca online in modo da avere maggiori dettagli su quel soggetto.

Ma come fa questa tecnologia a riconoscere effettivamente se il soggetto è un volto celebre o meno?

Rekognition contiene, sulla sua piattaforma, un database predefinito di volti di celebrità costantemente aggiornato. Quando viene caricata un'immagine o un video, la funzione *Celebrity Recognition* confronta i volti presenti con quelli del database di celebrità. Se trova una corrispondenza, restituisce il nome della celebrità insieme a un punteggio di confidenza, che indica la probabilità che il volto riconosciuto corrisponda alla celebrità identificata.

3.2.6 Rilevamento dei dispositivi DPI

Amazon Rekognition, inoltre, è in grado di rilevare i dispositivi di protezione individuale (DPI) indossati dalle persone in un'immagine. Ciò è considerato uno strumento molto utile per migliorare le pratiche di sicurezza sul lavoro.

Per rilevare i dispositivi di protezione individuale in un'immagine, viene utilizzata l'API *DetectProtectiveEquipment* del servizio.

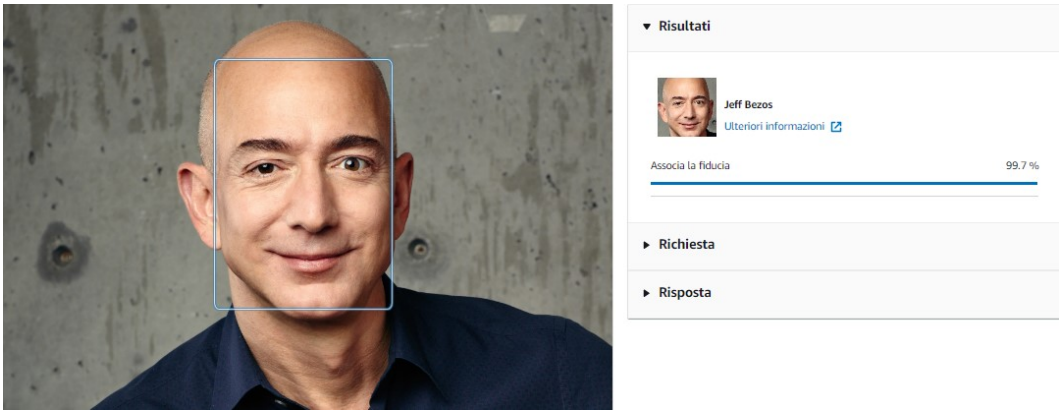


Figura 3.7: Funzione di riconoscimento di volti celebri

In particolare, per l'analisi, si fornisce l'immagine come input (come nella Figura 3.8) e la risposta, successivamente, restituisce una struttura *JSON* che include:

- l'elenco delle persone individuate nell'immagine;
- le parti del corpo dove sono indossati i DPI, come viso, testa, mano sinistra e mano destra;
- i tipi di DPI rilevati su ciascuna parte del corpo, tra cui protezioni per il viso, mani e copricapo;
- per ciascun DPI rilevato, è incluso un indicatore che specifica se il dispositivo copre adeguatamente la parte del corpo corrispondente;
- i riquadri di delimitazione che mostrano la posizione delle persone e degli oggetti DPI identificati nell'immagine.

Inoltre, è possibile richiedere un riepilogo degli articoli DPI e delle persone rilevati nell'immagine (Figura 3.9).

Figura 3.8: Persona che indossa i dispositivi di protezione individuale (DPI)

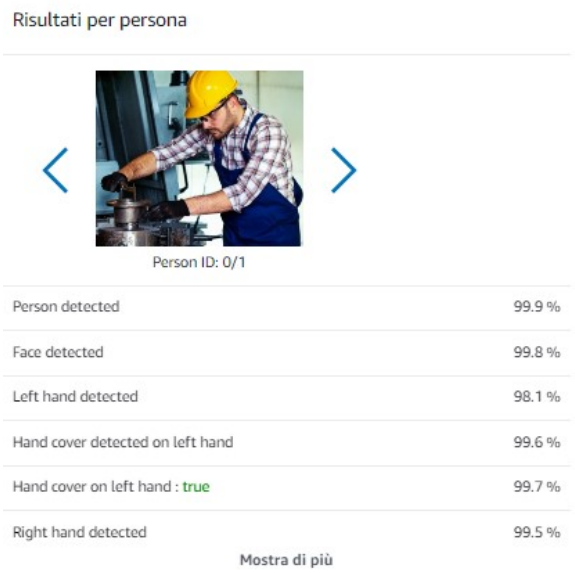


Figura 3.9: Risultati del rilevamento dei DPI

3.3 Applicazione di Amazon Rekognition

Nell'ambito della funzione Custom Labels, abbiamo addestrato un modello secondo le regole di Amazon Rekognition. Analizziamo in dettaglio tutti i passaggi elencati di seguito:

- creazione del progetto;
- creazione dei dataset: addestramento e test;
- Image Labeling;
- Model training;
- evaluate.

3.3.1 Creazione del progetto

Il primo step per l'addestramento di un modello è la creazione del progetto sulla console di AWS.

Il sistema di Amazon Web Service richiede una root in cui inserire e salvare le immagini per l'analisi con Custom Labels; ciò è possibile selezionando una delle seguenti possibilità:

- importare le immagini dal bucket S3;
- caricare le immagini dal proprio computer;
- copiare un set esistente di dati di etichette personalizzate di Amazon Rekognition;
- importare immagini etichettate da Amazon SageMaker Ground Truth.

In questo caso si è scelto di usare l'importazione delle immagini in una cartella del bucket S3 (Figura 3.10) di Amazon, chiamata "orchidea", in quanto essa è considerata una scelta molto più efficiente e ordinata per l'analisi e la selezione delle immagini per il dataset.

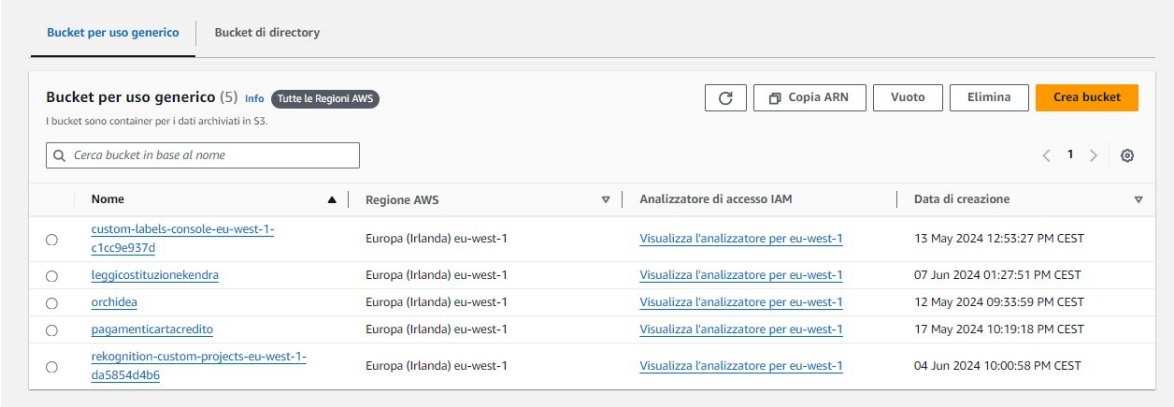


Figura 3.10: Bucket S3 contenente i dataset per Amazon Rekognition

3.3.2 Creazione dei dataset

La creazione del progetto e la selezione della fonte da cui estrarre le risorse corrispondono, in sostanza, alla fase di creazione del dataset. In dettaglio, questo passaggio consiste nella selezione di una grande quantità di immagini catturate da ogni angolazione possibile e con illuminazione più eterogenea possibile. Queste verranno, poi, suddivise in due dataset diversi, uno per l'addestramento, così da poter allenare il modello in modo efficiente, ed uno per il test, utile per valutare le prestazioni del modello.

Il dataset creato contiene circa 200 immagini, raffiguranti varie specie di orchidee (Figura 3.11)

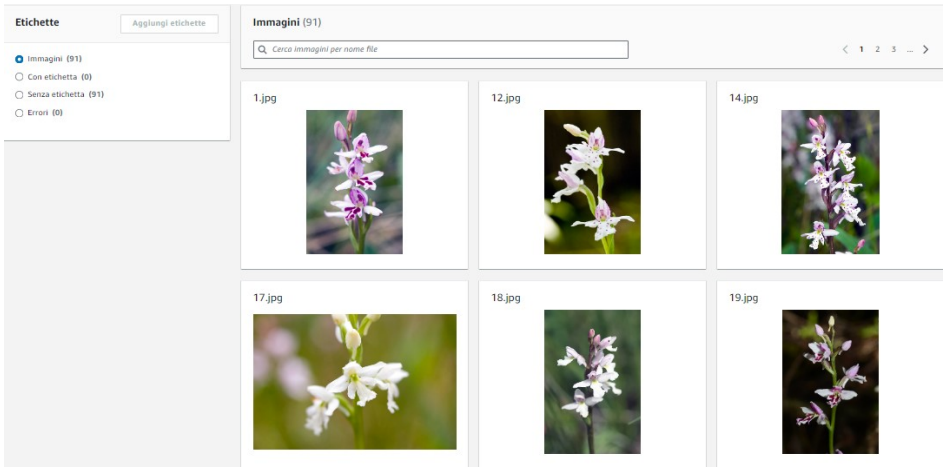


Figura 3.11: Cattura del dataset senza etichette

3.3.3 Image Labeling

In seguito alla creazione del dataset, si procede all'etichettamento manuale delle immagini di addestramento e di test, con un processo chiamato *Labeling*. In particolare, in questo passaggio si fornisce una breve descrizione indicando le caratteristiche, come, ad esempio, il colore e la specie (Figura 3.12) di ogni singola immagine. In questo modo, il server può allenarsi a riconoscere le diverse tipologie di orchidee facilmente.

Figura 3.12: Etichette personalizzate per l'analisi con Custom Labels

3.3.4 Model training

Dopo aver completato l'etichettatura delle immagini, possiamo procedere con l'addestramento del modello per riconoscere le etichette definite in precedenza. Selezionando l'opzione "train model", il sistema avvierà il processo (Figura 3.13), che può richiedere un tempo variabile, solitamente compreso tra 30 minuti e 24 ore. La durata dipende da diversi fattori, tra cui il numero di immagini, la quantità di etichette e la qualità delle immagini stesse.

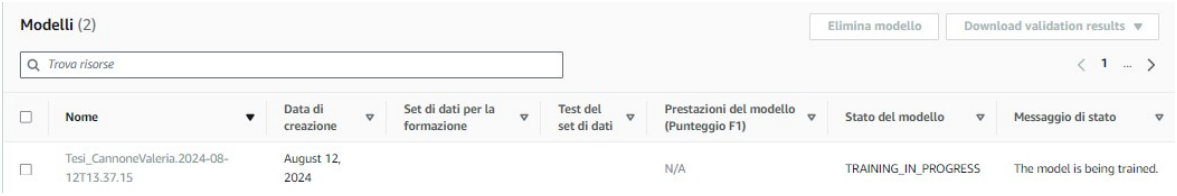


Figura 3.13: Processo di training del modello Custom Labels

Le etichette definite per questo modello si suddividono a seconda delle specie di orchidee, ovvero:

- Orchidea Anguloa Uniflora;
- Orchidea Caleana Major;
- Orchidea Calopogon Tuberosus;
- Orchidea Galearis Rotundifolia;
- Orchidea Pectellis Radiata;
- Orchidea Peristeria;
- Orchidea Phalaenopsis.

Appena il modello avrà completato l'allenamento, ci verrà inviata una notifica con il messaggio "TRAINING_COMPLETED" (Figura 3.14) e potremmo controllare la valutazione che ci viene offerta da AWS, contenente le statistiche riguardanti il processo di addestramento del nostro programma.

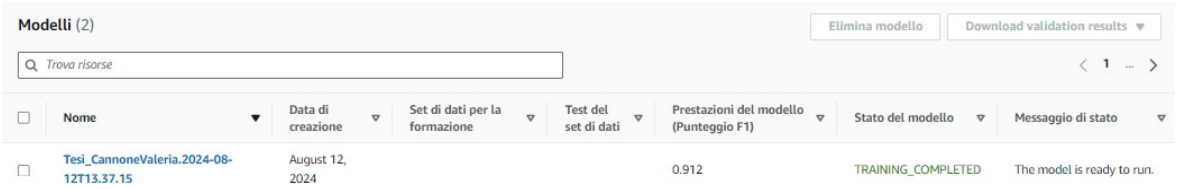


Figura 3.14: Valutazione generale ottenuta del modello analizzato

3.3.5 Evaluate

Cliccando sul nome del nostro progetto, verremo reindirizzati alla pagina di valutazione del nostro modello (Figura 3.15) in cui possiamo visionare le statistiche riguardanti il processo di training del nostro programma come, ad esempio:

- Il tempo impiegato per completare l'operazione (in questo caso, 17 minuti).
- La Precision media, ovvero la frazione di previsioni corrette (veri positivi) su tutte le previsioni del modello (veri e falsi positivi) (in questo caso, 0.965).
- Il Recall medio, ovvero la frazione delle etichette del dataset di test che sono state predette correttamente, (in questo caso, 0.935).
- L' F1-score, ovvero una misura aggregata che tiene conto sia della Precision che del Recall su tutte le etichette (in questo caso, 0.948).

Inoltre, nella schermata di valutazione, è possibile visionare le prestazioni sopra menzionate per ogni etichetta inserita; in aggiunta, il sistema ci propone il valore di "soglia ipotizzata per un'etichetta", ovvero il valore al di sopra del quale una predizione viene considerata vera o falsa positiva.

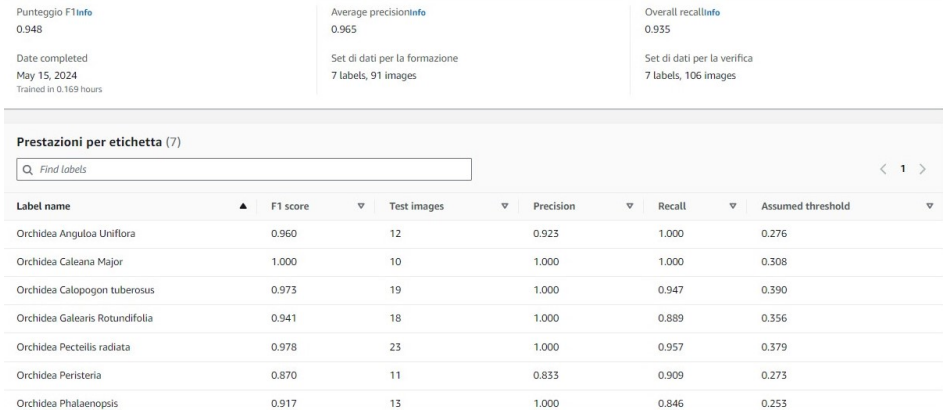


Figura 3.15: Cattura della schermata di valutazione di una Custom Label

Successivamente, è possibile analizzare, caso per caso, i risultati ottenuti sul dataset di test del nostro modello. Cliccando su "View test result" all'interno della schermata, è possibile esaminare ogni immagine fornita nella raccolta di test e controllare se l'algoritmo ha interpretato bene il contenuto della foto (Figura 3.16).

La galleria visualizza i risultati di ciascuna immagine esaminata. Ogni immagine di prova ha un'etichetta considerata come la "verità di base", ovvero quella inserita da noi dopo aver svolto tutte le osservazioni e gli studi necessari. Durante il test, questa etichetta viene confrontata con quella prevista dal modello, generando tre possibili scenari:

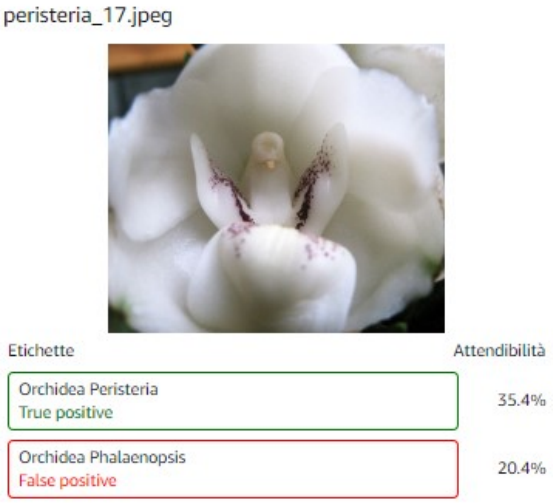


Figura 3.16: Cattura del risultato ottenuto dall'algoritmo su una singola immagine

- *Vero positivo*: quando l'etichetta prevista corrisponde a quella di verità di base dell'immagine.
- *Falso positivo*: quando l'etichetta prevista dal modello non coincide con la verità di base dell'immagine.
- *Falso negativo*: quando il modello non riesce a predire un'etichetta presente nella verità di base.

Il sistema fornisce anche un punteggio di fiducia, che quantifica la certezza del modello nella sua previsione per ciascuna etichetta (un punteggio più alto indica una maggiore fiducia).

3.4 Casi di studio

Amazon Rekognition offre diverse funzionalità utili per diversi settori della vita reale. In questo sottocapitolo analizzeremo il modo in cui alcune aziende usufruiscono di questo servizio AWS.

3.4.1 Social Media

Nell'ambito dei social media, Rekogniton è ritenuto uno strumento utile per proteggere gli utenti da contenuti inappropriati sulle piattaforme social e sui servizi che includono la condivisione di foto e video, i giochi online, lo streaming video e le app di appuntamenti online.

Pinterest

Pinterest è una piattaforma basata sulla condivisione di immagini, video e fotografie. Gli utenti possono esplorare, salvare e condividere i multimedia come "Pin" su bacheche di ispirazione digitali personalizzate.

La piattaforma, nata nel cloud di Amazon Web Services, ha trovato la soluzione ideale per scalare l'elaborazione, l'archiviazione e l'analisi dei suoi dati in costante crescita. Ciò ha

permesso all'azienda di ridurre la gestione dell'infrastruttura, concentrandosi maggiormente sull'innovazione.

Pinterest sfrutta il Machine Learning per migliorare il suo strumento di ricerca visiva, *Pinterest Lens*, che ora è in grado di riconoscere oltre 2.5 miliardi di oggetti e abbinarli ai prodotti. Inoltre, Pinterest ha una piattaforma di etichettatura preesistente, che ha integrato con servizi Amazon, come *Amazon Mechanical Turk*, un servizio di crowdsourcing online che consente ai programmatori di coordinare l'impiego di persone per svolgere compiti che i computer non riescono ancora ad eseguire. Recentemente, ha potenziato i suoi servizi di ricerca grazie all'integrazione di *SageMaker Ground Truth*, un servizio di etichettatura intuitivo, per ampliare le funzionalità della propria piattaforma e supportare le attività di etichettatura dei rettangoli di selezione.

La scoperta visiva su questa piattaforma avviene in diversi modi: tramite la ricerca, dove l'utente può utilizzare parte di un'immagine o l'intera immagine per cercare altri contenuti simili, oppure tramite personalizzazioni, le quali tengono conto delle interazioni passate degli utenti sulla piattaforma generando query specifiche per trovare contenuti rilevanti.

Snapchat

Snapchat è un'app multimediale per smartphone e tablet, la cui caratteristica principale è consentire agli utenti della propria rete di inviare gli "snap", ovvero messaggi di testo, foto e video visualizzabili solo per 24 ore.

Inizialmente, essa era basata su un'architettura monolitica nativa del cloud. Con la crescita della popolarità dell'applicazione, Snapchat è passata ad un'architettura a microservizi su Amazon Web Services (AWS) per migliorare la scalabilità, ottimizzare la disponibilità, minimizzare la latenza e ridurre i costi.

Successivamente, la piattaforma ha sperimentato l'uso di Amazon Rekognition per migliorare diverse funzionalità basate su immagini, come il filtraggio facciale e la moderazione dei contenuti. Grazie al riconoscimento facciale, l'app ha reso più accurato e dinamico l'uso dei filtri facciali sugli snap degli utenti. Inoltre, la moderazione dei contenuti, alimentata da Rekognition, ha aiutato a mantenere la piattaforma sicura, rilevando e bloccando contenuti potenzialmente inappropriati o violenti.

L'analisi delle immagini integrata sull'app ha anche permesso di migliorare la categorizzazione e la ricerca dei contenuti condivisi, offrendo suggerimenti più pertinenti agli utenti.

3.4.2 Videogiochi e sport

Nel mondo dei videogiochi e dello sport, l'adozione di tecnologie avanzate sta rivoluzionando il modo in cui gli utenti interagiscono con le piattaforme e il modo con cui vengono gestiti gli eventi sportivi. Tra queste tecnologie, Amazon Rekognition emerge come uno strumento innovativo che apporta significativi benefici in questi settori.

Nel contesto dei videogiochi, esso è utilizzato per migliorare la sicurezza della piattaforma, garantire l'integrità degli utenti e personalizzare l'esperienza di gioco. Nello sport, Rekognition contribuisce ad ottimizzare la gestione delle immagini e a migliorare l'accuratezza nella catalogazione e identificazione dei partecipanti agli eventi.

FanFight

FanFight è una piattaforma di Fantasy Sports, in cui gli utenti possono creare squadre virtuali e competere in vari sport, come cricket, calcio, basket, e altro ancora.

La piattaforma utilizza tecnologie avanzate per migliorare l'esperienza utente, ed uno dei casi studio più interessanti riguarda appunto l'integrazione di Amazon Rekognition.

L'obiettivo era quello di migliorare la sicurezza e l'integrità della piattaforma affrontando problemi come la creazione di account falsi, di attività fraudolente e comportamenti inappropriati tra gli utenti. Infatti, l'uso del servizio AWS ha consentito a FanFight di introdurre i seguenti aspetti:

- *Verifica dell'identità dell'utente*: FanFight utilizza Amazon Rekognition per confrontare le foto caricate dagli utenti con i loro documenti di identità ufficiali. Questa funzionalità aiuta a prevenire la creazione di account falsi o multipli da parte dello stesso utente.
- *Rilevamento di contenuti inappropriati*: il sistema analizza le immagini caricate dagli utenti per rilevare contenuti inappropriati, come immagini violente o sessualmente esplicite, e ne impedisce la pubblicazione.
- *Miglioramento della sicurezza*: Amazon Rekognition è anche utilizzato per monitorare comportamenti sospetti o fraudolenti, come l'uso di identità rubate, aiutando a bloccare tali attività prima che causino problemi.
- *Personalizzazione dell'esperienza utent*e: attraverso l'analisi delle immagini e dei video, FanFight può migliorare la personalizzazione dell'esperienza utente, ad esempio suggerendo contenuti o competizioni basate sugli interessi visibili nelle immagini del profilo dell'utente.

Dopo l'integrazione della tecnologia AWS sulla piattaforma, FanFight ha visto un miglioramento significativo nella sicurezza e nella fiducia degli utenti, con una riduzione delle attività fraudolente e una maggiore soddisfazione degli utenti.

Sportograf

Sportograf è una piattaforma specializzata nella fotografia sportiva, fornendo servizi di foto e video per eventi sportivi in tutto il mondo. Sportograf è nota per catturare immagini di alta qualità di partecipanti a gare di corsa, ciclismo, triathlon, e altre competizioni sportive, offrendo poi la possibilità di acquistare queste foto online.

Esso utilizza Amazon Rekognition per migliorare l'efficienza e la precisione nella gestione delle immagini, offrendo un servizio più rapido e accurato ai partecipanti degli eventi sportivi.

Uno degli usi principali del servizio per Spotograf consiste nel riconoscimento del numero di pettorale dei partecipanti. In particolare, durante un evento, date le migliaia di foto che vengono scattate, ogni partecipante è identificato principalmente tramite il numero di pettorale. Amazon Rekognition permette di automatizzare il processo di identificazione dei numeri di pettorale nelle immagini, associando le foto ai partecipanti in modo rapido e preciso.

In alcuni casi, Sportograf utilizza il riconoscimento facciale per migliorare l'accuratezza nella categorizzazione delle immagini, specialmente quando i numeri di pettorale negli sportivi non sono visibili o sono parzialmente coperti. Per questo motivo, l'uso di Amazon Rekognition permette di rilevare e confrontare i volti nelle immagini con quelli registrati nei profili utente, aumentando le possibilità di trovare le foto corrette.

Sebbene Sportograf operi principalmente in un contesto sportivo, l'uso del servizio AWS per il rilevamento di contenuti inappropriati potrebbe essere rilevante, ad esempio per garantire che le immagini pubblicate e vendute siano conformi agli standard di qualità e contenuto.

Con migliaia di immagini generate per ogni evento, Amazon Rekognition aiuta ad organizzare ed indicizzare le foto, rendendo più facile per i partecipanti trovare le proprie immagini. Il servizio può etichettare le immagini con metadati rilevanti, come la posizione del partecipante nel percorso o il momento specifico della gara.

L'integrazione di Amazon Rekognition, quindi, migliora l'esperienza utente complessiva su Sportograf, permettendo ai partecipanti di trovare le loro foto più velocemente e con maggiore precisione, grazie alla combinazione di queste funzionalità.

3.4.3 E-commerce e pubblicità

La capacità di Rekognition nell'analizzare e comprendere le immagini e i video in modo dettagliato offre opportunità significative per migliorare le operazioni aziendali e ottimizzare le strategie di marketing. Proprio per questo motivo, aziende come Amazon e Coca-Cola, utilizzano questo servizio.

Amazon

Amazon è, sicuramente, il leader nell'utilizzo di Rekognition. Esso è il "creatore" di questo servizio; infatti, offre le sue API ad aziende ed enti privati, fornendo loro numerose funzionalità di riconoscimento ed analisi dei media.

Inoltre, utilizza questa tecnologia all'interno della sua piattaforma per migliorare l'esperienza di shopping visivo. I clienti possono caricare foto dei prodotti desiderati e la tecnologia aiuta a trovare articoli simili o correlati nel catalogo dell'e-commerce.

Coca-Cola

Coca-Cola, invece, sfrutta Amazon Rekognition per analizzare e classificare le immagini e i video nelle sue campagne pubblicitarie, assicurando che i contenuti rispettino gli standard aziendali e riducendo il rischio di associazioni con contesti indesiderati o controversi.

La tecnologia di riconoscimento visivo consente di personalizzare gli annunci pubblicitari, identificando le preferenze visive del pubblico e creando campagne più rilevanti e coinvolgenti in base agli interessi e ai comportamenti degli utenti.

In questo modo, Rekognition aiuta Coca-Cola a ottimizzare le sue strategie di marketing, a proteggere il brand e a prendere decisioni basate su dati concreti.

3.4.4 Sicurezza

Nel settore della sicurezza, Amazon Rekognition viene impiegato per diversi scopi, tra cui il monitoraggio delle telecamere di sorveglianza, l'autenticazione degli accessi e la prevenzione delle frodi. Alcune delle aziende che traggono vantaggio dalle funzionalità di sicurezza offerte da questa tecnologia sono Software Colombia e Abode Systems.

Software Colombia

Software Colombia è una società di sviluppo software di alto livello con sede a Bogotà, in Colombia, che fornisce soluzioni tecnologiche all'avanguardia a livello globale. Alcune di queste sono l'apprendimento automatico (ML), l'Intelligenza Artificiale (AI), lo sviluppo software, lo sviluppo di app mobile, lo sviluppo web, il Cloud Computing e i Big Data. L'azienda opera in altri ambiti, come la sanità, la finanza, la logistica e l'istruzione.

Per affrontare le difficoltà di gestione dell'identità digitale e migliorare la sicurezza delle transazioni, Software Colombia ha sviluppato una soluzione chiamata *eLogic Biometrics*. Questa innovativa piattaforma utilizza Amazon Rekognition per fornire un sistema di riconoscimento facciale e autenticazione biometrica avanzata.

In particolare, *eLogic Biometrics* è progettata per mitigare gli attacchi di spoofing dell'identità, riducendo il rischio di frodi del 95%, grazie alla verifica avanzata dei tratti del viso. Essa consente, inoltre, di ottimizzare i processi di onboarding, riducendo il tempo di verifica e autenticazione degli utenti del 92% e migliorando, di conseguenza, l'esperienza utente.

Abode Systems

Abode Systems è un'azienda specializzata in soluzioni di sicurezza domestica e automazione che offre una vasta gamma di prodotti, come telecamere di sorveglianza, sensori di movimento, allarmi e sistemi di automazione domestica.

Amazon Rekognition Streaming Video Events rappresenta uno strumento particolarmente utile per aziende come Abode Systems poiché consente il rilevamento rapido di persone, animali domestici e pacchi. Ciò aiuta a ridurre il numero di notifiche relative a movimenti irrilevanti, garantendo, al contempo, una maggiore sicurezza per i clienti.

3.4.5 Media e intrattenimento

Nel settore dei media e dell'intrattenimento, l'innovazione tecnologica gioca un ruolo cruciale nel migliorare l'esperienza degli utenti e nel rendere più efficienti le operazioni aziendali. Due esempi significativi dell'applicazione di Amazon Rekognition in questo ambito sono rappresentati da K-STAR Group e C-SPAN.

C-SPAN

C-SPAN è una rete televisiva via cavo statunitense che offre agli spettatori notizie e attualità sugli avvenimenti di tipo politico che si svolgono negli stati della federazione.

Gli archivi di C-SPAN documentano 24 ore su 24 e 7 giorni su 7 le attività di otto reti C-SPAN, inclusa la copertura in diretta di Senato e Camera dei Rappresentanti, oltre a flussi esclusivi online non trasmessi. I programmi sono accuratamente indicizzati, offrendo una risorsa pubblica unica per l'istruzione, la ricerca, la revisione e l'accesso dei cittadini.

L'uso di Amazon Rekognition è stato particolarmente rilevante in questo; infatti, ha permesso al team di C-SPAN di indicizzare il doppio dei contenuti rispetto a quelli indicizzati in precedenza passando, da 3500 ore all'anno a 7500 ore all'anno e ottenendo una percentuale di indicizzazione del 100%.

La soluzione carica screenshot presi a intervalli di sei secondi da tutti gli otto feed di C-SPAN e li confronta con una raccolta di immagini, rilevando, così, il soggetto che si sta cercando. Inoltre, per ridurre i costi, viene utilizzato il rilevamento degli scatti per eliminare l'analisi delle immagini duplicate. Ciò ha portato, quindi, numerosi vantaggi, quali la riduzione del lavoro da 1 ora a 20 minuti, il riconoscimento facciale altamente accurato e il caricamento di 97.000 immagini in meno di due ore.

KSTAR Group

K-STAR Group è una società sudcoreana che opera nel settore dell'intrattenimento, specializzata nella fornitura di servizi di biglietteria e pagamento per concerti ed eventi live. Per affrontare il problema delle lunghe file ai concerti, dove i partecipanti devono aspettare per mostrare e far convalidare il biglietto cartaceo all'ingresso, K-STAR ha deciso di impiegare

Amazon Rekognition per sviluppare il servizio Face Ticket. Questo sistema permette ai partecipanti di verificare rapidamente il loro acquisto o di scansionare il biglietto cartaceo all'ingresso, riducendo significativamente le attese.

In sintesi, i casi d'uso analizzati dimostrano la versatilità e l'efficacia del modello in vari contesti applicativi, evidenziando come l'integrazione di tali soluzioni possa portare significativi vantaggi operativi e strategici. La capacità di adattarsi a diverse esigenze e settori rende il servizio AWS una risorsa preziosa, capace di trasformare i dati in informazioni utili per prendere decisioni più consapevoli e migliorare i processi aziendali.

CAPITOLO 4

Esperienze sul rilevamento delle frodi

Le frodi stanno diventando sempre più frequenti, diffondendosi in vari ambiti, dal settore finanziario a quello digitale. Questa crescente minaccia richiede una solida comprensione delle modalità con cui esse possono manifestarsi e delle strategie efficaci per contrastarle. In questo capitolo, esamineremo i diversi comportamenti fraudolenti e gli approcci migliori per identificarli e prevenirli. Verranno analizzati i metodi tradizionali e le tecnologie più avanzate per la gestione delle frodi, con particolare attenzione agli strumenti digitali. Infine, verrà illustrato come utilizzare Amazon Fraud Detector di AWS per creare modelli predittivi di rilevamento delle frodi, offrendo una risorsa preziosa sia per le aziende che per i privati nella lotta contro questo fenomeno.

4.1 Introduzione al rilevamento delle frodi

Il rilevamento delle frodi rappresenta una componente cruciale nella protezione delle risorse finanziarie e reputazionali di organizzazioni e individui.

Con l'aumento della digitalizzazione e la sofisticazione delle tecniche fraudolente, le aziende e le istituzioni devono implementare strategie efficaci per identificare e prevenire attività fraudolente. Infatti, negli ultimi decenni, le tecniche di rilevamento delle frodi sono evolute notevolmente, passando da metodi tradizionali a soluzioni avanzate basate su Intelligenza Artificiale e Machine Learning. Di conseguenza, anche i criminali si adattano rapidamente alle misure di sicurezza, richiedendo soluzioni sempre più sofisticate e rendendo le sfide nel rilevamento delle frodi molto più complesse.

La frode, o truffa, è un comportamento consistente in artifizi o raggiri per indurre altre persone in errore al fine di conseguire illeciti profitti. Queste attività illecite possono causare perdite economiche significative, danneggiare la reputazione e generare costi aggiuntivi per le aziende.

Esistono varie tipologie di frodi, ovvero:

- *frodi online*: phishing, frodi con carta di credito, frodi sui siti di e-commerce;
- *frodi informatiche*: attacchi hacker, phishing, malware, manipolazione di dati;
- *frodi finanziarie*: Insider Trading, manipolazione dei mercati;
- *frodi aziendali*: furto di identità, appropriazione indebita di fondi, frodi di bilancio, corruzione;
- *frodi nei servizi*: frodi assicurative, frodi sanitarie;

- *frodi alimentari*: frodi sanitarie, frodi commerciali;
- *frodi fiscali*: evasione.

Analizziamo i casi specifici e le tecniche di rilevamento per combattere le minacce.

4.1.1 Frodi online

Le frodi online si riferiscono ad attività fraudolente che avvengono su Internet e che coinvolgono la manipolazione dei dati, delle identità o delle transazioni.

I principali rischi che si possono riscontrare sono:

- *phishing*: attacchi in cui i truffatori inviano email o messaggi falsi per indurre le persone a fornire informazioni sensibili, come password o dati di carte di credito;
- *frodi con carta di credito*: esse consistono nell'utilizzo illecito delle informazioni della carta di credito di un utente per effettuare acquisti online senza il suo consenso;
- *frodi sui siti di e-commerce*: i truffatori possono vendere beni inesistenti o contraffatti, o manipolare le transazioni per ottenere pagamenti senza consegnare il prodotto.

4.1.2 Frodi informatiche

La frode informatica è un reato disciplinato dall'articolo 640 ter del codice penale, introdotto dalla legge n. 547/1993. Essa comprende qualsiasi tipo di attività fraudolenta che utilizza la tecnologia informatica come mezzo principale causando i seguenti problemi:

- alterazione del funzionamento del sistema informatico;
- intervento non autorizzato su dati, informazioni o programmi;
- manipolazione delle informazioni contenute nel sistema.

Esempi di questa tipologia di frode includono:

- *attacchi hacker*: accesso non autorizzato a sistemi informatici per rubare, alterare o distruggere dati, oppure per compromettere il funzionamento dei sistemi;
- *malware*: software malevolo progettato per infiltrarsi nei sistemi informatici e causare danni, rubare informazioni o prendere il controllo del dispositivo infettato;
- *manipolazione di dati*: quest'operazione consiste nell'alterare, falsificare o distorcere dati all'interno di un sistema informatico con l'intento di ingannare o trarre un vantaggio illegittimo.

4.1.3 Frodi finanziarie

Le frodi finanziarie comprendono una vasta gamma di pratiche illecite volte ad ottenere vantaggi economici attraverso la manipolazione delle informazioni finanziarie. Tra le principali vi sono:

- *Insider Trading*: esso consiste nell'acquisto o la vendita di titoli (come azioni o obbligazioni) da parte di persone che hanno accesso a informazioni riservate e non pubblicamente disponibili, che possono influenzare il prezzo di quei titoli.
- *manipolazione dei mercati*: pratiche che influenzano artificialmente il prezzo o il volume di titoli, attraverso azioni come la diffusione di false informazioni o il coordinamento di acquisti e vendite.

4.1.4 Frodi aziendali

Le frodi aziendali rappresentano atti fraudolenti compiuti all'interno di un'azienda, spesso da parte di dipendenti, dirigenti o terzi, a danno della stessa organizzazione.

Esempi di frodi aziendali includono:

- *furto di identità*: esso consiste nell'utilizzo illecito dell'identità di un'altra persona per ottenere vantaggi finanziari o per accedere a risorse aziendali;
- *appropriazione indebita di fondi*: ciò consiste nell'uso non autorizzato di fondi aziendali per fini personali da parte di un dipendente o di un dirigente;
- *frode di bilancio*: essa si occupa della manipolazione delle informazioni finanziarie di un'azienda per nascondere perdite, aumentare profitti o ingannare investitori e regolatori;
- *corruzione*: atti di corruzione includono il pagamento o la ricezione di somme per ottenere vantaggi aziendali indebiti.

4.1.5 Frodi nei servizi

Le frodi nei servizi riguardano, maggiormente, l'inganno e la manipolazione all'interno di settori come quello assicurativo e quello sanitario; esempi di frodi nei servizi includono:

- *frodi assicurative*: esse includono pratiche fraudolente per ottenere pagamenti indebiti dalle compagnie di assicurazione, come presentare falsi sinistri o aumentare l'entità del danno;
- *frodi sanitarie*: esse includono comportamenti fraudolenti nel settore sanitario, come la fatturazione di servizi medici mai erogati, l'uso di identità rubate per ottenere servizi sanitari o la vendita di farmaci contraffatti.

4.1.6 Frodi fiscali

Le frodi fiscali comprendono l'evasione e altre pratiche illecite finalizzate a ridurre, oppure evitare, il pagamento delle imposte. Una delle frodi fiscali più frequenti è l'evasione fiscale, ovvero il tentativo di ridurre illegalmente l'obbligo fiscale attraverso la sottodichiarazione dei redditi, la falsa dichiarazione dei costi o l'occultamento al fisco di beni e redditi.

4.1.7 Frodi alimentari

Le frodi alimentari si verificano quando cibo o bevande vengono alterati o etichettati in modo ingannevole per ottenere un guadagno economico illecito. Esse si suddividono in frodi sanitarie e frodi commerciali, entrambe con l'obiettivo di ingannare i consumatori e ottenere benefici indebiti.

Le frodi sanitarie riguardano l'uso di ingredienti pericolosi, contaminati o non conformi agli standard di sicurezza, che possono compromettere la salute dei consumatori. Ad esempio, esse includono l'aggiunta di sostanze tossiche o la presenza di contaminanti nel prodotto alimentare. Questi tipi di frode rappresentano un grave rischio per la salute pubblica e sono spesso soggetti a severe sanzioni legali.

Le frodi commerciali, invece, riguardano pratiche ingannevoli che influenzano la qualità, la composizione o l'origine dei prodotti alimentari, senza necessariamente compromettere direttamente la salute. Esempi di queste frodi includono la sostituzione di ingredienti di

alta qualità con quelli di qualità inferiore o la falsa dichiarazione sull'origine geografica del prodotto. Questi inganni mirano a sfruttare la reputazione di ingredienti pregiati o a sfruttare l'attrattiva di determinate origini per ottenere un vantaggio economico.

4.1.8 Tecniche di rilevamento delle frodi

Due componenti essenziali per combattere le frodi sono l'individuazione e la prevenzione delle stesse. L'individuazione consiste nel riconoscere comportamenti sospetti o anomali, mentre la prevenzione si riferisce alle misure da adottare per evitare questi inconvenienti.

Sebbene gli approcci classici di rilevamento delle frodi siano ancora ampiamente utilizzati e rappresentino un buon punto di partenza per un'organizzazione, si sta sperimentando un nuovo metodo basato sui dati o su analisi statistiche per tre ragioni in particolare:

- *Precisione*: i sistemi basati sui dati migliorano la rilevazione delle frodi riducendo al minimo il numero di falsi positivi (cioè i casi non fraudolenti che vengono erroneamente identificati come tali) e massimizzando il numero di frodi reali scoperte tra quelle esaminate. Ciò consente di ottimizzare l'uso delle risorse limitate destinate alle ispezioni.
- *Efficienza operativa*: essa riguarda la capacità di un sistema di identificare le frodi in modo rapido e accurato attraverso l'uso di analisi automatizzate su grandi volumi di informazioni. Per massimizzare l'efficienza operativa, i sistemi di rilevamento delle frodi basati sui dati possono essere integrati con altri sistemi aziendali, come la gestione dei rischi, il CRM (Customer Relationship Management) o i sistemi di pagamento.
- *Efficienza dei costi*: l'efficienza sui costi si riferisce alla capacità di ridurre le spese associate all'identificazione e alla gestione delle frodi, massimizzando, allo stesso tempo, l'efficacia del processo. I sistemi basati sui dati migliorano i costi riducendo il numero di falsi positivi, automatizzando il processo di rilevamento e prevenendo le perdite derivanti dagli attacchi.

Prima di procedere oltre, nel seguito presenteremo una panoramica sulle altre tecniche di rilevamento delle frodi.

Rilevamento basato su regole

Il rilevamento basato su regole prevede l'applicazione di una serie di regole predefinite per identificare attività sospette. Tali regole sono sviluppate sulla base di esperienze passate e conoscenze esperte. Ad esempio, una regola potrebbe bloccare transazioni superiori a un certo importo o provenienti da regioni ad alto rischio. Questo metodo è semplice da implementare, ma può risultare limitato nel rilevare schemi di frode complessi o nuovi.

Questa tecnica di rilevamento è stata sperimentata con Amazon Fraud Detector, come vedremo in dettaglio più avanti.

Analisi comparativa

L'analisi comparativa confronta il comportamento di un'entità, come un cliente o un dipendente, con quello di un gruppo di riferimento simile. Discrepanze significative possono indicare attività fraudolente. Ad esempio, un cliente che spende improvvisamente molto più della media dei clienti appartenenti alla stessa categoria potrebbe essere segnalato per ulteriori verifiche. Questa tecnica è utile per identificare anomalie rispetto ai comportamenti tipici.

Rilevamento basato sulle anomalie

Il rilevamento basato su anomalie identifica comportamenti che si discostano dalla norma. Sebbene questa tecnica possa essere integrata con il rilevamento basato sui dati, è particolarmente efficace per individuare attività inusuali che potrebbero indicare una frode. Un esempio potrebbe essere una transazione effettuata in un paese in cui il cliente non ha mai viaggiato. Questo metodo è particolarmente utile per rilevare frodi non ancora note.

Verifiche manuali e investigazioni

Le verifiche manuali e le investigazioni consistono nell'esaminare manualmente le transazioni o le attività sospette da parte di un team di esperti. Questo approccio è intensivo in termini di tempo e risorse, ma è per di più molto efficace per analizzare casi complessi che i sistemi automatizzati non riescono a risolvere. Ad esempio, una revisione manuale delle fatture può rilevare discrepanze nei prezzi o nelle quantità.

Segnalazioni di whistleblower

Un whistleblower è una persona che segnala, pubblicamente o segretamente, attività illegali o comportamenti scorretti verificatisi all'interno di un'organizzazione.

Nel contesto del rilevamento delle frodi, le segnalazioni di un whistleblower rappresentano una fonte preziosa di informazioni. Dipendenti, clienti o altri stakeholder possono denunciare attività sospette attraverso canali dedicati, spesso in anonimato. Queste segnalazioni forniscono un metodo diretto per individuare frodi interne che, altrimenti, potrebbero passare inosservate.

Controlli interni e revisione periodica

I controlli interni e le revisioni periodiche prevedono l'implementazione di processi e politiche interne per prevenire o rilevare frodi attraverso un controllo continuo e sistematico. Ad esempio, la revisione trimestrale dei bilanci aziendali può identificare discrepanze indicative di frodi. Questo approccio aiuta a mantenere un controllo costante sull'integrità delle operazioni aziendali.

Rilevamento comportamentale

Il rilevamento comportamentale analizza i modelli di comportamento degli utenti, come il modo in cui interagiscono con un sito web o utilizzano i propri account. Cambiamenti improvvisi o comportamenti atipici possono essere segnali di attività fraudolente. Ad esempio, un utente che accede in modo insolito al proprio account bancario potrebbe essere segnalato per ulteriori controlli.

Forensic accounting

Il forensic accounting consiste nell'uso di tecniche contabili avanzate per investigare e analizzare transazioni finanziarie sospette, spesso in contesti legali. Questo approccio è particolarmente utile per scoprire frodi complesse nascoste nei bilanci aziendali. Analisi dettagliate dei registri finanziari possono rivelare discrepanze significative che indicano attività fraudolente.

Monitoraggio delle transazioni in tempo reale

Il monitoraggio delle transazioni in tempo reale consente di rilevare attività sospette mentre avvengono, permettendo di bloccare immediatamente le frodi. Questo approccio è particolarmente efficace in contesti con alta frequenza di transazioni, come i sistemi di pagamento. Ad esempio, una transazione sospetta può essere bloccata sul nascere se rilevata in tempo reale.

Collaborazione con le forze dell'ordine e le agenzie di regolamentazione

La collaborazione con le forze dell'ordine e le agenzie di regolamentazione può migliorare l'efficacia del rilevamento delle frodi su larga scala. Le organizzazioni possono scambiare informazioni e risorse con queste entità per identificare e contrastare frodi che coinvolgono più parti o che operano su scala globale. Questo approccio aiuta a creare un ambiente di controllo più robusto e integrato.

Questi approcci tradizionali possono essere potenziati grazie all'integrazione di metodi più sofisticati e avanzati, come le tecniche di *apprendimento supervisionato* e di *analisi predittiva*.

Questi metodi, utilizzando i dati storici, consentono di costruire modelli capaci di distinguere i comportamenti legittimi dai comportamenti sospetti, individuando anche i cosiddetti "allarmi silenziosi", ovvero quelle tracce che i truffatori non riescono a cancellare.

Tuttavia, le tecniche di apprendimento supervisionato e l'analisi predittiva presentano alcuni limiti. In particolare, esse dipendono fortemente dalla disponibilità di dati storici etichettati. Ciò significa che possono avere difficoltà nel riconoscere nuove tipologie di frode che utilizzano dei metodi ancora non documentati.

Per affrontare questo problema, l'analisi descrittiva può risultare più efficace, poiché è in grado di rilevare deviazioni rispetto al comportamento normale, anche se il metodo fraudolento è completamente nuovo.

Un'ulteriore innovazione nella lotta contro le frodi è rappresentata dall'analisi dei social network, che consente di studiare il comportamento fraudolento all'interno di una rete di soggetti interrelati. Questa tecnica, pur essendo relativamente recente, si è dimostrata molto potente nel rafforzare i sistemi di rilevazione delle frodi.

In questo contesto, diventa evidente come l'integrazione di metodi supervisionati, non supervisionati e l'analisi dei social network possa creare un sistema di rilevazione delle frodi altamente efficace e adattabile alle nuove minacce.

4.2 Sicurezza e prevenzione delle frodi

La sicurezza e la prevenzione delle frodi richiedono un approccio strutturato che combina misure proattive e reattive per identificare, ridurre e gestire i rischi. Esplorare strategie efficaci in questo ambito è essenziale per garantire la protezione dell'organizzazione e prevenire perdite significative.

Un'analisi delle metodologie tradizionali per contrastare furti, frodi e appropriazioni indebite da parte dei dipendenti rappresenta un passo fondamentale per elaborare un programma di prevenzione e controllo delle frodi efficaci.

Gli approcci classici adottati all'interno delle organizzazioni che analizzeremo sono i seguenti:

- *Approccio direttivo*: l'approccio direttivo è conflittuale e autorevole. Esso si concentra sulla prevenzione delle frodi attraverso l'implementazione di politiche, procedure e controlli interni, chiari e rigorosi, implementando sistemi di controllo, come, ad esempio, l'uso di software per monitorare le attività.

- *Approccio preventivo*: l'approccio preventivo alla gestione delle frodi si concentra sulla riduzione delle opportunità di comportamento fraudolento prima che questo si verifichi. Ciò avviene attraverso l'implementazione di misure e controlli progettati per prevenire la frode. Per esempio, le organizzazioni possono utilizzare strumenti particolari, come i controlli sui precedenti penali e le informazioni creditizie, per identificare potenziali truffatori durante il processo di assunzione. Creare un ambiente di controllo ben strutturato e promuovere una cultura aziendale etica sono elementi chiave di questa tipologia di approccio.

- *Approccio osservativo*: l'approccio osservativo si basa sulla sorveglianza continua delle operazioni e dei comportamenti dei dipendenti per identificare segnali di potenziali frodi. Questo metodo mira a rilevare comportamenti sospetti prima che si trasformino in azioni fraudolente vere e proprie. In sintesi, si basa sul monitoraggio e nella supervisione delle azioni effettuate dagli utenti.

- *Approccio investigativo*: l'approccio investigativo viene messo in atto quando c'è il sospetto o l'evidenza di una frode. Questo metodo implica una ricerca approfondita per raccogliere prove, determinare la natura e l'entità della frode e identificare i responsabili. Esso include le indagini forensi, le interviste e le collaborazioni con le autorità esterne.

- *Approccio assicurativo*: l'approccio assicurativo si basa sull'utilizzo di polizze assicurative per proteggere l'organizzazione dalle perdite finanziarie derivanti da frodi ed altre attività illecite. Questo approccio non si concentra tanto sulla prevenzione diretta delle frodi, quanto sulla mitigazione del loro impatto economico.

Questi sei approcci, combinati tra di loro, forniscono un quadro completo per prevenire, rilevare ed affrontare le frodi all'interno di un'organizzazione.

Al di fuori del contesto aziendale, la prevenzione fa riferimento ad una serie di strategie, pratiche e misure che possono essere adottate anche in diversi ambiti della società, come il settore pubblico, il mondo digitale, il commercio, e persino a livello personale. Ecco una panoramica generale dei principali approcci alla prevenzione delle frodi nei vari contesti:

- *Approccio educativo e sensibilizzante*: in questa tipologia di approccio è richiesta una formazione approfondita sulle frodi e sulle truffe attraverso campagne di sensibilizzazione pubblica per il loro riconoscimento. Inoltre, sarebbe consigliato educare le persone sul piano finanziario per ridurre la vulnerabilità alle frodi derivanti da investimenti e prestiti accattivanti per gli utenti.

- *Approccio tecnologico*: alla base dell'approccio tecnologico c'è la sicurezza digitale che riguarda l'uso di tecnologie, come la crittografia, l'autenticazione a più fattori e i software antivirus, utili a proteggere dati sensibili e a prevenire frodi digitali. Analogamente, l'implementazione di sistemi avanzati di monitoraggio, spesso alimentati dall'Intelligenza Artificiale, può aiutare a rilevare comportamenti anomali e sospetti, sia nelle transazioni finanziarie che nelle attività online, permettendo un intervento rapido.

- *Approccio legale e normativo*: esso è gestito da autorità governative e organizzazioni di regolamentazione che svolgono un ruolo cruciale nel monitorare le attività economiche e commerciali per prevenire le frodi.

- *Approccio comportamentale*: esso si basa su uno screening dei comportamenti in contesti come quello bancario o assicurativo, analizzando i comportamenti e i profili di rischio delle persone in modo da identificare potenziali frodatori prima che possano agire.

- *Approccio assicurativo*: come nelle organizzazioni, anche gli individui e le piccole imprese possono proteggersi dalle frodi attraverso polizze assicurative.
- *Approccio collaborativo*: nell'approccio collaborativo prevale la cooperazione intersettoriale, quindi tra diverse entità come, ad esempio, le agenzie governative, le istituzioni finanziarie, le aziende e il pubblico. È possibile, anche, utilizzare piattaforme di segnalazione e di allerta, ovvero sistemi in cui le persone possono segnalare tentativi di frode o di truffe già avvenute.

4.2.1 Tecnologie per la sicurezza e la prevenzione delle frodi

Le tecnologie più utilizzate negli ultimi anni per combattere le crescenti frodi sono l'Intelligenza Artificiale, l'Autenticazione Multifattoriale (MFA), la crittografia e le Blockchain. Analizziamole di seguito.

Intelligenza Artificiale

L'IA utilizza diverse tecnologie avanzate, quali il Machine Learning, la Computer Vision e l'OCR (Optical Character Recognition), per analizzare grandi quantità di dati ed identificare degli schemi sospetti.

Ad esempio, gli algoritmi di *Machine Learning* analizzano modelli di transazioni per riconoscere attività sospette. Lo vedremo più avanti con il servizio di AWS, Amazon Fraud Detector.

La *Computer Vision* è utile per analizzare immagini e video; essa, quindi, permette di contrastare le frodi riconoscendo documenti contraffatti, rilevando comportamenti sospetti tramite le telecamere di sicurezza e verificando l'autenticità dei prodotti.

L'*OCR*, ovvero il riconoscimento ottico dei caratteri, è una tecnologia che permette di convertire immagini contenenti testo scritto, come documenti scansionati o foto di testi stampati, in testo digitale modificabile. L'OCR migliora l'efficienza nel rilevamento delle frodi riducendo l'intervento umano e permettendo di analizzare grandi quantità di dati in modo rapido e accurato.

Autenticazione Multifattoriale (MFA)

L'Autenticazione Multifattoriale è una tecnologia di sicurezza che utilizza diversi fattori di autenticazione per la verifica dell'identità quando gli utenti accedono ai servizi online. Essa comprende le password, i PIN, i servizi di impronte digitali o di geolocalizzazione, le Smart Card o le password monouso (OTP).

Questo approccio rende molto più difficile, per i malintenzionati, ottenere l'accesso non autorizzato, poiché devono superare più livelli di sicurezza. Infatti, l'MFA è una potente difesa contro le frodi, particolarmente efficace nell'era digitale in cui le minacce sono in continua evoluzione.

Crittografia

La crittografia protegge le informazioni sensibili attraverso la trasformazione dei dati in un formato illeggibile, che può essere decifrato solo da coloro che possiedono la chiave di decrittazione appropriata.

Questo metodo è utile per le transazioni online e per le comunicazioni interne ad un'organizzazione. In particolare, la crittografia *SSL/TLS* può essere usata nei pagamenti digitali e nell'e-commerce e protegge le informazioni di pagamento dagli attacchi tra il cliente e il

venditore. Invece, nelle organizzazioni, le comunicazioni vengono crittografate per prevenire accessi non autorizzati da parte dei dipendenti o hacker.

Blockchain

La tecnologia Blockchain è una tecnologia emergente che offre vantaggi significativi, come l'integrità dei dati, la trasparenza e la sicurezza. Tuttavia, presenta anche rischi legati alle frodi informatiche, come l'*attacco a 51%* e l'*attacco del doppio speso*. Nel primo caso l'aggressore ottiene il controllo della maggioranza della potenza di calcolo di una blockchain, potendo così manipolare le transazioni e creare monete false; nell'attacco del doppio speso, poichè le transazioni vengono salvate su dei blocchi, l'attaccante spende gli stessi fondi due volte, creando un blocco falso.

Per questo motivo, per difendersi dalle frodi informatiche sulla blockchain, sarebbe opportuno utilizzare una blockchain sicura, aggiornare il software frequentemente ed utilizzare una buona sicurezza informatica, come, ad esempio, password forti e autenticazione a due fattori, per proteggere i propri account.

4.2.2 Il Data Mining e l'analisi predittiva

Due strumenti che si sono diffusi con lo sviluppo delle tecnologie informatiche per la prevenzione delle frodi sono il Data Mining e l'analisi predittiva.

Il Data Mining è l'insieme delle metodologie utilizzate per estrarre informazioni da una grande quantità di dati ed è in grado di identificare associazioni nascoste tramite le tecniche di apprendimento automatico.

In seguito alla raccolta delle informazioni, il processo prevede la costruzione di un modello di analisi predittiva. Lo scopo è quello di classificare un soggetto non ancora verificato come potenziale frodatore assegnandogli uno score (ovvero un punteggio) che fornisce un'indicazione sulla sua propensione al comportamento identificato come anomalo.

I dati vengono prima analizzati con le tecniche di Data Mining per identificare le caratteristiche rilevanti e poi vengono utilizzati per alimentare i modelli predittivi.

L'uso congiunto di Data Mining e analisi predittiva nella prevenzione delle frodi permette di rilevare comportamenti sospetti con maggiore precisione, proteggendo, così, le organizzazioni e i loro clienti da perdite finanziarie e di reputazione.

Per quanto riguarda la sicurezza, attraverso il Data Mining è possibile analizzare i log di sistema, i dati di rete e i comportamenti degli utenti per identificare, tramite i dati storici, attività anomale che potrebbero indicare tentativi di frode o attacchi informatici. Qualsiasi deviazione significativa da questi modelli può essere segnalata come potenziale minaccia.

Utilizzando i modelli predittivi, è possibile prevedere la probabilità di futuri attacchi informatici, come phishing, il ransomware o attacchi DDoS, basandosi su dati storici e pattern di attacchi passati.

4.2.3 Sicurezza mobile e IoT

Wikipedia definisce i dispositivi di sicurezza elettronici IoT come strumenti connessi ad una rete tramite Internet, progettati per offrire misure di sicurezza. Questi dispositivi, controllabili da remoto tramite app mobili, interfacce web o software proprietario, includono spesso funzionalità come il monitoraggio dei video, il rilevamento delle intrusioni, gli avvisi automatici e l'automazione intelligente.

La sicurezza nell'IoT richiede l'adozione di tecnologie come la crittografia, per proteggere le informazioni, l'autenticazione dei dispositivi e dei server, il controllo degli accessi e l'implementazione di sistemi di rilevamento delle intrusioni, al fine di minimizzare i rischi.

Essa è cruciale per la protezione dei dati sensibili e la salvaguardia della privacy, evitando gli accessi non autorizzati.

I dispositivi connessi ad Internet, come quelli per la casa intelligente, la tecnologia indossabile e le macchine industriali, sono particolarmente vulnerabili rispetto alle frodi. In particolare, i dispositivi che raccolgono e trasmettono i dati personali, i video, gli audio o i dati finanziari sono obiettivi particolarmente appetibili per gli hacker, aumentando il rischio di attacchi. Per questo motivo, è essenziale che le organizzazioni e gli utenti adottino misure di sicurezza rigorose per proteggere tali informazioni e garantire la loro privacy e la loro sicurezza.

Analogamente, i dispositivi e i sistemi IoT nel settore sanitario e domestico, se non adeguatamente protetti, presentano significativi rischi di sicurezza. Ad esempio, i sistemi di diagnostica per immagini, che archiviano dati sensibili sui pazienti, sono vulnerabili a cyberattacchi che possono interrompere i servizi ospedalieri e facilitare il furto d'identità o le frodi finanziarie. Infatti, i sistemi di monitoraggio dei pazienti, se non aggiornati e crittografati, sono esposti a violazioni dei dati e ransomware, mettendo a rischio la privacy e la sicurezza dei pazienti.

Anche i telefoni IP e i dispositivi di gestione dell'energia, come i contatori intelligenti, sono soggetti agli attacchi. Infatti, attraverso questi strumenti, gli hacker possono intercettare comunicazioni o manipolare dati energetici, portando a fatturazioni errate o interruzioni di servizio.

Infine, le telecamere di sicurezza, spesso dotate di protezioni minime, rappresentano un ulteriore punto di vulnerabilità, con il rischio di accessi non autorizzati e potenziali violazioni della privacy. In tutti questi casi è essenziale adottare misure di sicurezza robuste e mantenere aggiornati i sistemi per mitigare i rischi.

4.3 Amazon Fraud Detector

Amazon Fraud Detector è un servizio di Machine Learning di AWS basato sull'apprendimento automatico che permette ai clienti di identificare potenziali attività fraudolente in modo che sia possibile rilevare le frodi online più velocemente. Tali attività includono le transazioni non autorizzate e la creazione di account falsi.

Amazon Fraud Detector consente di creare modelli personalizzati di rilevamento delle frodi, aggiungere la logica decisionale e le regole specifiche per interpretare le valutazioni delle frodi e, di conseguenza, assegnare risultati come "approvato" o "in attesa di revisione" come valutazione di ciascuna frode. Ciò è possibile anche in assenza di conoscenze teoriche sul rilevamento delle frodi. Analizzeremo, nello specifico, queste proprietà in seguito.

I principali casi d'uso di Amazon Fraud Detector sono:

- l'identificazione di pagamenti sospetti online;
- il rilevamento delle frodi su un nuovo account;
- il servizio online e l'abuso dei programmi fedeltà;
- il rilevamento dell'acquisizione di account.

4.3.1 Benefici

Amazon Fraud Detector offre una serie di vantaggi utili a velocizzare il rilevamento delle frodi. Essi sono di seguito specificati.

Creazione automatizzata di modelli di frode

Essendo completamente automatizzati e specializzati per soddisfare le esigenze aziendali, i modelli di AWS per il rilevamento delle frodi permettono di rinunciare a molti dei passaggi associati alla loro creazione e al loro addestramento. Questi passaggi includono la convalida e l'arricchimento dei dati, la preparazione delle variabili, la selezione dell'algoritmo, la definizione degli iperparametri e l'implementazione del modello.

Infatti, per creare un modello di rilevamento delle frodi utilizzando Amazon Fraud Detector, è necessario caricare il set di dati delle operazioni finanziarie dell'azienda e selezionare il tipo di modello. In seguito, Amazon Fraud Detector trova automaticamente l'algoritmo di rilevamento delle frodi più adatto al caso d'uso e crea il modello.

Modelli di frode che si evolvono e apprendono

Amazon Fraud Detector è in grado di aggiornare automaticamente i modelli di rilevamento delle frodi, utilizzando informazioni come l'età dell'account, l'intervallo di tempo dall'ultima attività e il numero totale di attività registrate. Grazie a questi parametri, il modello è in grado di distinguere tra utenti affidabili, che effettuano transazioni regolari, e comportamenti sospetti tipici dei truffatori. Questo meccanismo aiuta a mantenere elevate le prestazioni del modello anche tra le diverse sessioni di riaddestramento.

Visualizzazione delle prestazioni del modello di frode

Dopo l'addestramento del modello con i dati forniti, è possibile valutare il punteggio delle prestazioni del modello, il grafico di distribuzione del punteggio, la matrice di confusione, la tabella delle soglie e tutti gli input forniti, classificati in base al loro impatto sulle prestazioni del modello. Questi strumenti permettono di comprendere meglio il funzionamento del modello e, se necessario, di apportare modifiche per ottimizzarne le prestazioni.

Previsione delle frodi

Il servizio AWS genera previsioni utilizzando la logica di previsione con i dati associati all'attività. Esso effettua previsioni per una singola attività in tempo reale oppure previsioni di frode offline per un set di attività.

Visualizzazione della spiegazione della previsione delle frodi

Amazon Fraud Detector fornisce spiegazioni dettagliate delle previsioni come parte del processo di rilevamento delle frodi. Queste spiegazioni offrono dettagli su come ciascun elemento di dati, utilizzato nell'addestramento del modello, ha influenzato il punteggio di previsione delle frodi. Le spiegazioni vengono presentate attraverso strumenti visivi, come grafici e tabelle, e permettono di visualizzare l'impatto di ogni dato sui risultati delle previsioni. Le informazioni ottenute possono essere utilizzate per analizzare i modelli di comportamento fraudolento relativi ai set di dati ed identificare eventuali distorsioni. Inoltre, esse sono utili per evidenziare i principali indicatori di rischio durante le indagini manuali sulle frodi, aiutando l'analista ad individuare le cause principali dei falsi positivi.

Azioni basate su regole

Per ottenere un modello di addestramento avanzato, è utile definire delle regole ricavate dallo studio del set di dati.

4.3.2 Come funziona Amazon Fraud Detector

Amazon Fraud Detector è un servizio che crea un modello di Machine Learning personalizzato per identificare potenziali attività fraudolente in un'azienda. Come primo step è necessario fornire dei dettagli specifici sul proprio caso d'uso aziendale. In base a questi dettagli, Amazon Fraud Detector suggerisce un tipo di modello da utilizzare ed indica quali elementi di dati storici fornire. Questi dati sono poi utilizzati per addestrare automaticamente un modello personalizzato.

Durante il processo di addestramento, viene scelto un algoritmo di Machine Learning specifico per il proprio caso d'uso. Questo processo include la convalida e la manipolazione dei dati per ottimizzare le prestazioni del modello. Al termine dell'addestramento, il servizio AWS produce dei punteggi e delle metriche da utilizzare per valutare l'efficacia del modello. Se necessario, bisogna modificare i dati utilizzati per l'addestramento e ripetere il processo per migliorare il modello.

Una volta creato e addestrato il modello, è necessario configurare delle regole decisionali, le quali indicano allo stesso come interpretare i dati aziendali e quali azioni intraprendere in base ai risultati. Queste azioni possono includere l'approvazione, la revisione o la classificazione del rischio (alto, medio, basso).

Il rilevatore è un contenitore che integra il modello e le regole configurate. Dopo averlo creato, testato e distribuito, esso può essere utilizzato nel proprio ambiente di produzione per rilevare frodi nelle applicazioni aziendali. Durante la valutazione, il modello confronta i dati in arrivo con quelli storici, applicando algoritmi di Machine Learning e le regole definite per determinare i risultati. Studiamo nello specifico l'applicazione di questi passaggi.

4.4 Applicazione di Amazon Fraud Detector

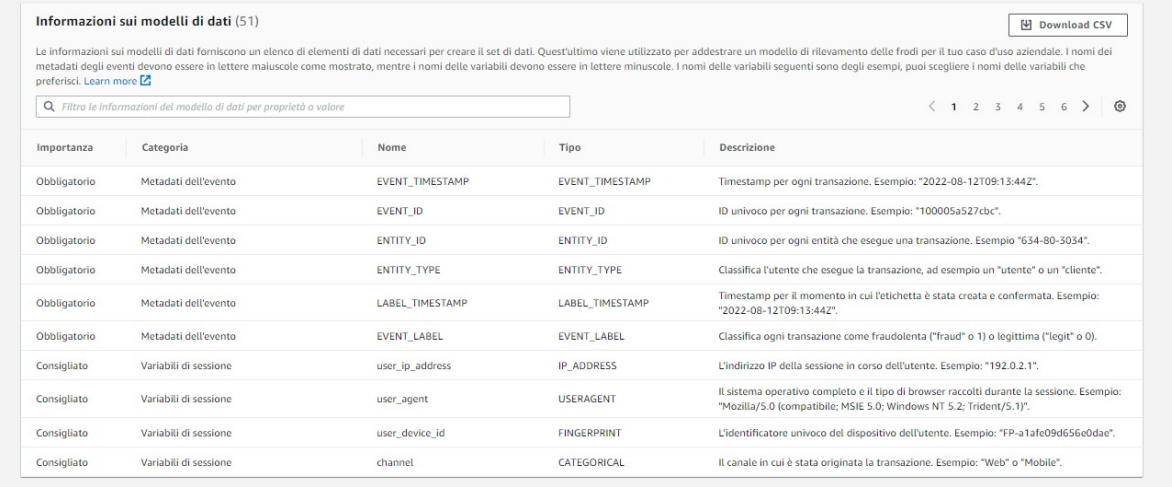
La creazione di un modello per il rilevamento di attività fraudolente in Amazon Fraud Detector richiede una serie di prerequisiti, tra cui la definzione dei dati di accesso dell'account utente AWS in un file `.CSV` codificato nel formato UTF-8.

Inoltre, è necessario fornire determinati metadati di evento e variabili di evento nella riga di intestazione del proprio file `.CSV` (necessariamente la prima riga del file) utilizzando solo lettere maiuscole.

I metadati richiesti per l'evento sono i seguenti:

- `EVENT_ID`: esso è l'identificatore univoco per l'evento di accesso;
- `ENTITY_TYPE`: esso indica l'entità che esegue l'evento di accesso, ad esempio un commerciante o un cliente;
- `ENTITY_ID`: esso rappresenta l'identificatore dell'entità che esegue l'evento di accesso;
- `EVENT_TIMESTAMP`: esso è un metadato dell'evento che specifica l'ora in cui quest'ultimo si è verificato. Il formato del timestamp deve essere nello standard ISO 8601 in UTC, ovvero del tipo yyyy-mm-ddThh:mm:ssZ;
- `EVENT_LABEL` (facoltativo): esso rappresenta un'etichetta che classifica l'evento come fraudolento o legittimo. È possibile usare qualsiasi etichetta, come *fraud, legit, 1 o 0.*

Ogni riga del file rappresenta un evento. Nel nostro caso è stato utilizzato il modello dei dati per le transazioni online (Figura 4.1):



Informazioni sui modelli di dati (51)

Download CSV

Le informazioni sui modelli di dati forniscono un elenco di elementi di dati necessari per creare il set di dati. Quest'ultimo viene utilizzato per addestrare un modello di rilevamento delle frodi per il tuo caso d'uso aziendale. I nomi dei metadati degli eventi devono essere in lettere maiuscole come mostrato, mentre i nomi delle variabili devono essere in lettere minuscole. I nomi delle variabili seguenti sono degli esempi, puoi scegliere i nomi delle variabili che preferisci. Learn more

Filtra le informazioni del modello di dati per proprietà o valore

1 2 3 4 5 6

Importanza	Categoria	Nome	Tipo	Descrizione
Obbligatorio	Metadati dell'evento	EVENT_TIMESTAMP	EVENT_TIMESTAMP	Timestamp per ogni transazione. Esempio: "2022-08-12T09:13:44Z".
Obbligatorio	Metadati dell'evento	EVENT_ID	EVENT_ID	ID univoco per ogni transazione. Esempio: "100005a527cbc".
Obbligatorio	Metadati dell'evento	ENTITY_ID	ENTITY_ID	ID univoco per ogni entità che esegue una transazione. Esempio "634-80-3034".
Obbligatorio	Metadati dell'evento	ENTITY_TYPE	ENTITY_TYPE	Classifica l'utente che esegue la transazione, ad esempio un "utente" o un "cliente".
Obbligatorio	Metadati dell'evento	LABEL_TIMESTAMP	LABEL_TIMESTAMP	Timestamp per il momento in cui l'etichetta è stata creata e confermata. Esempio: "2022-08-12T09:13:44Z".
Obbligatorio	Metadati dell'evento	EVENT_LABEL	EVENT_LABEL	Classifica ogni transazione come fraudolenta ("fraud" o 1) o legittima ("legit" o 0).
Consigliato	Variabili di sessione	user_ip_address	IP_ADDRESS	L'indirizzo IP della sessione in corso dell'utente. Esempio: "192.0.2.1".
Consigliato	Variabili di sessione	user_agent	USERAGENT	Il sistema operativo completo e il tipo di browser raccolti durante la sessione. Esempio: "Mozilla/5.0 (compatibile; MSIE 5.0; Windows NT 5.2; Trident/5.1)".
Consigliato	Variabili di sessione	user_device_id	FINGERPRINT	L'identificatore univoco del dispositivo dell'utente. Esempio: "FP-a1afe09d656e0dae".
Consigliato	Variabili di sessione	channel	CATEGORICAL	Il canale in cui è stata originata la transazione. Esempio: "Web" o "Mobile".

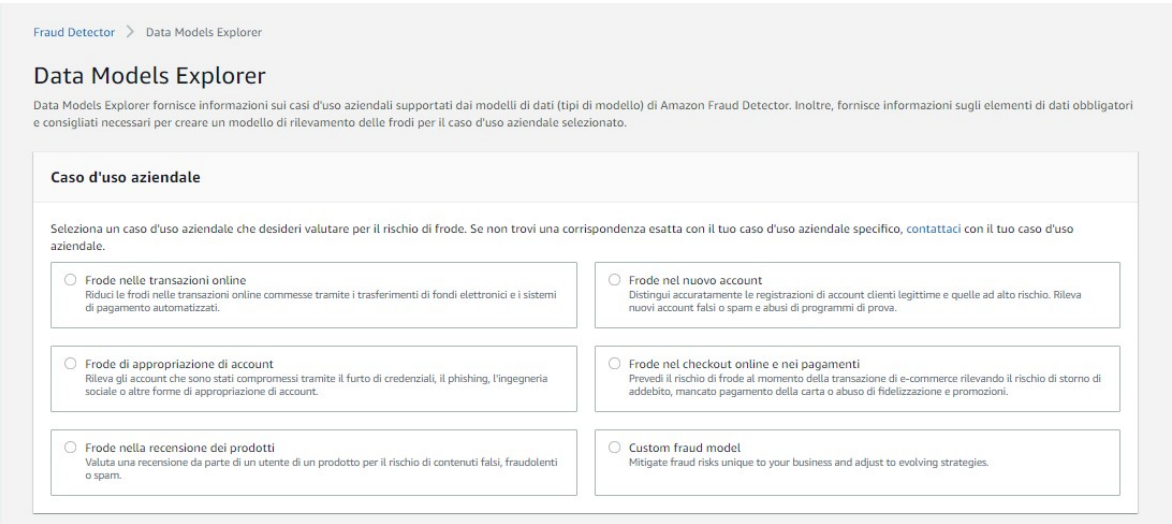
Figura 4.1: Modello dei dati per le transazioni online

4.4.1 Preparazione del set di dati

La preparazione del set di dati richiede che il file contenga almeno 1.500 entità (account utente individuali), ciascuna con almeno due eventi di accesso associati ed un set di dati che copre almeno 30 giorni di eventi di accesso. Quest'ultimo deve includere eventi definiti, in particolare, da due etichette personalizzate; nel nostro modello sono definite come fraud o legit, per indicare se la transazione è fraudolenta o leggittima.

4.4.2 Validazione dei dati

Amazon Fraud Detector effettua un controllo dei dati per constatare se effettivamente vengono rispettati tutti i requisiti di dimensione e formato. Se il set di dati non supera la convalida, non viene creato il modello ed è possibile consultare il Data Model Explorer (Figura 4.2), ovvero uno strumento nella console Amazon Fraud Detector che allinea il caso d'uso aziendale con il tipo di modello supportato dalla piattaforma. Esso fornisce un elenco di dati obbligatori, consigliati e facoltativi necessari per creare il proprio set di dati.



Fraud Detector > Data Models Explorer

Data Models Explorer

Data Models Explorer fornisce informazioni sui casi d'uso aziendali supportati dai modelli di dati (tipi di modello) di Amazon Fraud Detector. Inoltre, fornisce informazioni sugli elementi di dati obbligatori e consigliati necessari per creare un modello di rilevamento delle frodi per il caso d'uso aziendale selezionato.

Caso d'uso aziendale

Seleziona un caso d'uso aziendale che desideri valutare per il rischio di frode. Se non trovi una corrispondenza esatta con il tuo caso d'uso aziendale specifico, contattaci con il tuo caso d'uso aziendale.

- Frode nelle transazioni online — Riduci le frodi nelle transazioni online commesse tramite i trasferimenti di fondi elettronici e i sistemi di pagamento automatizzati.
- Frode nel nuovo account — Distingui accuratamente le registrazioni di account clienti legittime e quelle ad alto rischio. Rileva nuovi account falsi o spam e abusi di programmi di prova.
- Frode di appropriazione di account — Rileva gli account che sono stati compromessi tramite il furto di credenziali, il phishing, l'ingegneria sociale o altre forme di appropriazione di account.
- Frode nel checkout online e nei pagamenti — Prevedi il rischio di frode al momento della transazione di e-commerce rilevando il rischio di storno di addebito, mancato pagamento della carta o abuso di fidelizzazione e promozioni.
- Frode nella recensione dei prodotti — Valuta una recensione da parte di un utente di un prodotto per il rischio di contenuti falsi, fraudolenti o spam.
- Custom fraud model — Mitigate fraud risks unique to your business and adjust to evolving strategies.

Figura 4.2: Data Models Explorer

4.4.3 Definire l'entità, il tipo di evento e le variabili dell'evento

In questa sezione sono illustrate le componenti necessarie all'addestramento del modello e il modo in cui impostarle.

Definire l'entità

Un'entità rappresenta colui sta eseguendo l'evento. Nell'ambito di una previsione di frode, è possibile utilizzare l'ID entità per indicare l'entità specifica che ha eseguito l'evento. Le entità di esempio includono il cliente, l'esercente o l'account.

Per creare questa componente bisogna:

1. aprire la console di gestione AWS ed accedere al proprio account;
2. selezionare il servizio Amazon Fraud Detector, selezionare la sezione *Entità* nel menu di navigazione a sinistra, quindi cliccare su *Crea*;
3. nella pagina *Crea entità* (Figura 4.3), inserire *cliente* come nome del tipo di entità; facoltativamente, si può inserire una descrizione dell'entità;
4. selezionare *Crea entità*.

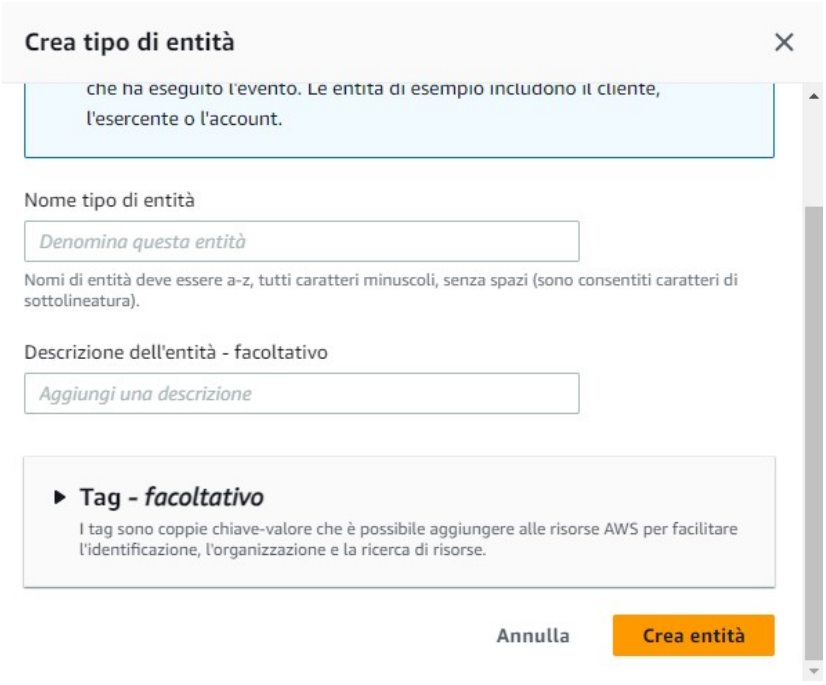


Figura 4.3: Creazione dell'elemento Entità sulla console AWS

Nello studio effettuato abbiamo inserito come entità il cliente, in quanto rappresenta il soggetto che ha effettuato le transazioni registrate nel file.

Definire l'evento e le variabili dell'evento

Un evento in Amazon Fraud Detector rappresenta un'azione specifica o un'interazione che si desidera monitorare e analizzare per individuare potenziali frodi. Un evento può essere, ad esempio, una transazione di pagamento, una registrazione di un nuovo account, o una richiesta di rimborso.

Le variabili dell' evento sono i dati specifici associati ad un evento e vengono utilizzati per l'analisi. Alcune di queste informazioni possono essere, ad esempio, l'importo della transazione, il paese del cliente, il metodo di pagamento utilizzato e l'indirizzo IP.

Le variabili dell' evento vengono utilizzate dai modelli di Amazon Fraud Detector per determinare se un evento è legittimo o fraudolento.

È possibile inserire queste componenti sempre attraverso la console di Amazon Fraud Detector nelle sezioni apposite, come nelle Figure 4.4 e 4.5.

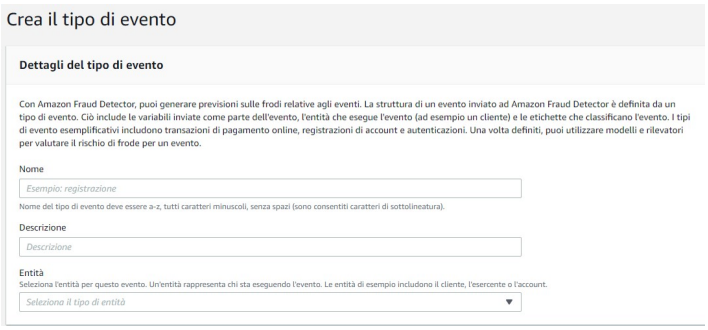


Figura 4.4: Creazione del tipo di evento

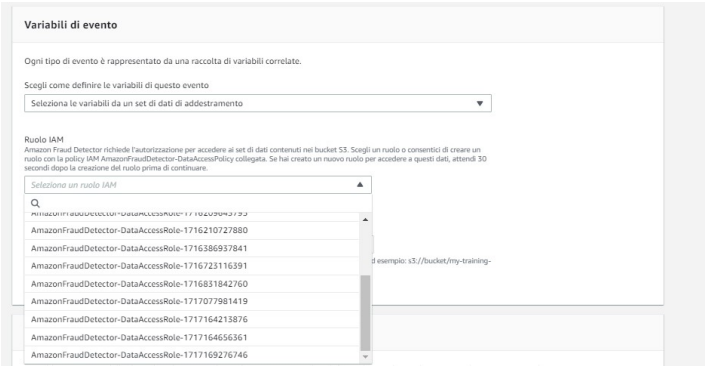


Figura 4.5: Selezione della variabile di evento e ruolo IAM

Nella nostra configurazione, abbiamo scelto di selezionare le variabili di evento da un set di dati di addestramento, ovvero un file `.CSV`, ma è possibile farlo anche dai modelli di Amazon SageMaker.

Successivamente, sarà necessario scegliere un *ruolo IAM (Identity and Access Management)*. È possibile creare un nuovo ruolo al momento oppure selezionare un ruolo esistente, già registrato nel servizio di gestione delle identità e degli accessi (IAM) di AWS.

Nello specifico, un ruolo IAM è un'entità IAM che definisce un set di autorizzazioni per fare richieste ai servizi AWS. Esso può essere assunto temporaneamente da entità diverse, permettendo ad esse di ottenere le autorizzazioni associate al ruolo.

Infine, sarà richiesta la posizione dei dati di addestramento sul servizio di storage di AWS, ovvero il Bucket S3, con il formato `s3://bucket/nome_file.csv`.

Amazon Fraud Detector, successivamente, estrae le intestazioni dal set di dati di addestramento inserito e crea una variabile per ogni intestazione.

4.4.4 Definizione delle etichette degli eventi

Le etichette (Figura 4.6) vengono utilizzate per classificare i singoli eventi come frode o legittimi. Esse sono facoltative per l'addestramento del modello perché il modello ATI

utilizza un approccio innovativo all'apprendimento non supervisionato. Però, è comunque consigliato includere il metadato EVENT_LABEL e fornire etichette per gli eventi di accesso, se disponibili.

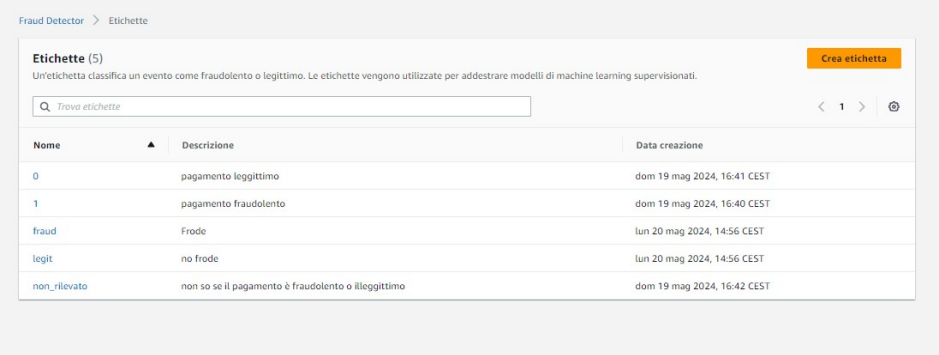


Figura 4.6: Etichette per la determinazione del tipo di evento

Le etichette utilizzate nel nostro modello sono *0* e *1*, perchè, nel file dei dati di addestramento, il metadato EVENT_LABEL definito per gli eventi contiene già queste etichette definite come variabili.

4.4.5 Valutare le prestazioni del modello e distribuirlo

Amazon Fraud Detector offre diversi strumenti per valutare le prestazioni del modello e per trovare il giusto equilibrio tra il rilevamento delle frodi e la minimizzazione del disagio per i clienti legittimi.

Una volta completato l'addestramento del modello, il sistema utilizza il 15% dei dati non utilizzati per convalidare le sue prestazioni (Figura 4.7), offrendo le seguenti metriche:

- *Anomaly Separation Index (ASI)*: essa riassume la capacità complessiva del modello di separare le attività anomale dal comportamento degli utenti previsto. Un modello senza potere di separabilità avrà il punteggio ASI più basso possibile, pari a 0,5. Al contrario, il modello con un potere di separabilità elevato avrà il punteggio ASI più alto possibile, pari a 1,0.
- *Tasso di sfida (CR)*: la soglia del punteggio indica la percentuale di eventi di accesso che il modello consiglierebbe di sfidare sotto forma di password monouso, di autenticazione a più fattori, di verifica dell'identità e di indagine.
- *Anomaly Discovery Rate (ADR)*: esso quantifica la percentuale di anomalie che il modello può rilevare rispetto la soglia di punteggio selezionata. Una soglia di punteggio più bassa aumenta la percentuale di anomalie catturate dal modello. Tuttavia, esso richiederebbe anche di mettere in discussione una percentuale più significativa di eventi di accesso, portando ad un maggiore disagio per il cliente.
- *ATO Discovery Rate (ATODR)*: esso quantifica la percentuale di eventi di compromissione dell'account che il modello può rilevare alla soglia di punteggio selezionata. Questa metrica è disponibile solo se nel set di dati analizzato sono presenti 50 o più entità con almeno un evento ATO etichettato.
- *Area Under the Curve (AUC)*: questo parametro riassume il TPR (Tasso di veri positivi, True Positive Rate in inglese) e l'FPR (Tasso di falsi positivi, False Positive Rate) su tutte le possibili soglie di punteggio del modello. Un modello senza potere predittivo ha un'AUC di 0,5, mentre un modello perfetto ha un punteggio di 1,0.

Quando il modello *transactions_model* ha terminato l'addestramento, è possibile visualizzare l'*Area under the curve (AUC)*. Un AUC dello 0,93-0,94% indica che il modello è quasi perfetto, quindi potrebbe essere usato come sistema di riferimento.

Scrivendo una regola utilizzando una soglia del punteggio del modello di 500, si riesce ad individuare una percentuale del 90.1% di tutti gli eventi fraudolenti, accettando, al contempo, il rischio che il 13.4% di eventi legittimi venga erroneamente etichetta come frode.

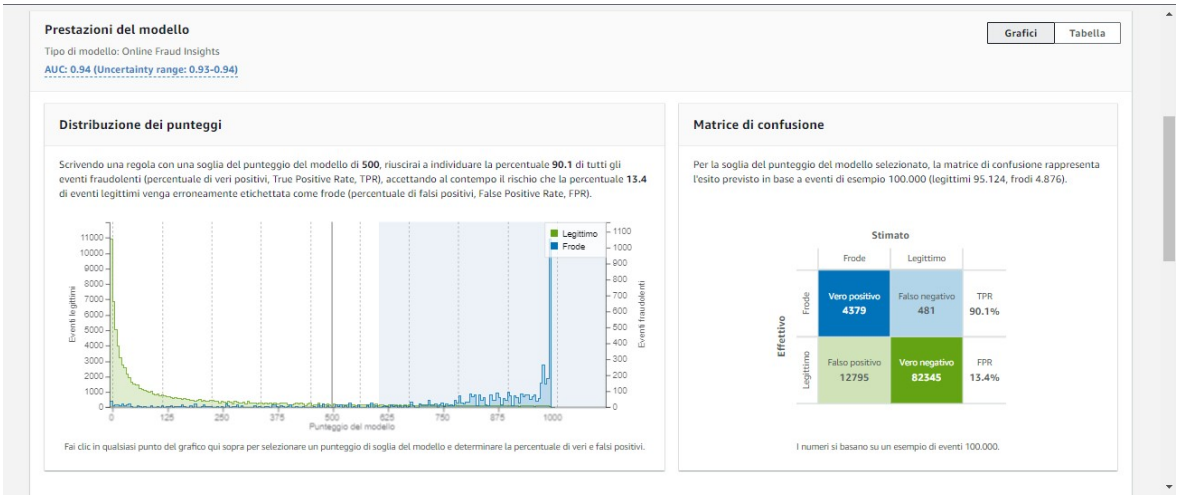


Figura 4.7: Prestazioni del modello

Un'altra metrica importante è l'importanza della variabile modello (Figura 4.8). Essa aiuta a capire come le diverse variabili influenzano le prestazioni del modello. Nel modello sono analizzate due tipi di variabili: quelle grezze e quelle aggregate. Le variabili grezze sono quelle che sono state definite in base al set di dati, mentre le variabili aggregate sono una combinazione di più variabili. In questo modello sono presenti solo variabili grezze.

Inoltre, la configurazione restituisce un valore di importanza per le singole variabili. Ciò indica in che percentuale quella determinata variabile di evento influisce sulla determinazione del modello.

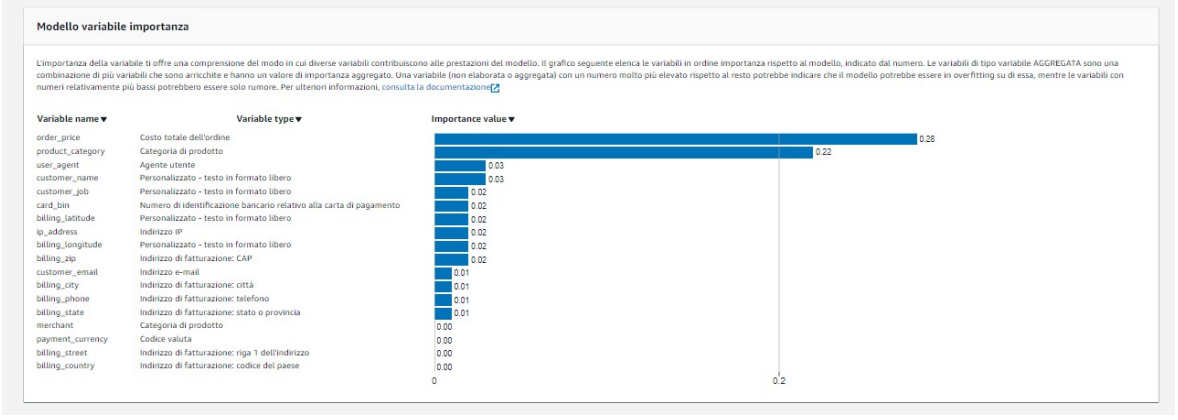


Figura 4.8: Modello delle variabili di importanza

Dopo aver esaminato le prestazioni del modello e dopo aver deciso quali soglie di punteggio del modello sono in linea con il modello aziendale, è possibile distribuire la versione del modello.

4.4.6 Creazione del rilevatore

Tramite l'apposita sezione, lo step seguente consiste nella creazione del rilevatore, il quale richiede un nome identificativo e il tipo di evento (creato in precedenza).

Il rilevatore utilizzerà un insieme di regole, definite al suo interno e descritte nella Tabella 4.1, che specificano le variabili da analizzare all'interno del file. Queste regole permettono di determinare con precisione se una transazione è fraudolenta o legittima. In particolare, esse sono descritte da formule *and* ed *or*, selezionate e create dopo un'accurato studio delle variabili del file `.CSV` e del loro legame.

nome_regola	esiti
valute_basso_rischio	no_frode
rileva_tramite_importo	rischio_frode
categorie_alto_rischio	rischio_frode
valute_alto_rischio	rischio_frode
categorie_basso_rischio	no_frode
legit_ordine	no_frode
numerocarta_valido	no_frode
pagamento_fraudolento	rischio_frode
pagamento_legittimo	no_frode
ordine_eventlabel	no_frode

Tabella 4.1: Regole associate al rilevatore per l'addestramento del modello

4.4.7 Avviare la formazione del modello

Dopo aver importato correttamente gli eventi, abbiamo tutto il necessario per avviare l'addestramento del modello.

In questo step è possibile scegliere se creare un proprio modello (come in Figura 4.9) oppure utilizzare uno dei modelli di Sage Maker.

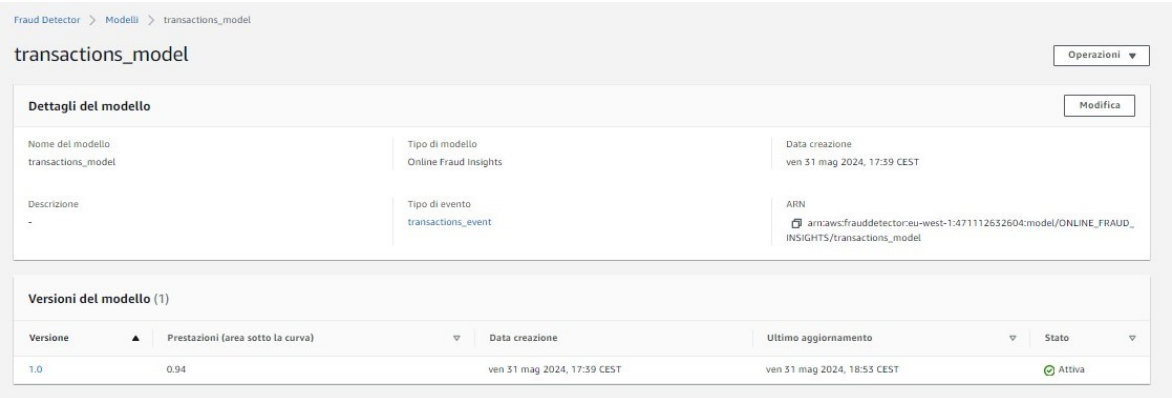


Figura 4.9: Modello di AWS Fraud Detector

Il modello utilizzato nel nostro addestramento è Online Fraud Insight.

Il modello utilizza un insieme di algoritmi di apprendimento automatico per trasformare, arricchire e classificare i dati relativi alle frodi. Durante il processo di addestramento, elementi come l'indirizzo IP e il numero BIN delle carte di credito, vengono arricchiti con informazioni aggiuntive, come la geolocalizzazione dell'indirizzo IP o la banca emittente della carta. Inoltre, il modello sfrutta modelli di Deep Learning basati sui pattern di frode identificati su Amazon e su AWS, e successivamente integrati in esso tramite un algoritmo di potenziamento dell'albero dei gradienti.

Dopo che il modello è stato addestrato, Amazon Fraud Detector fornisce un elenco classificato dell'impatto di ogni variabile sulle prestazioni, in modo da sapere se includere tale variabile nel futuro addestramento del modello. Se vengono fornite etichette, come nel caso che stiamo analizzando, Amazon Fraud Detector le utilizza per valutare e visualizzare le prestazioni del modello in termini di tasso di scoperta.

A questo punto, è possibile procedere con la creazione e l'addestramento definitivo del modello che richiederà circa 45 minuti. Quando il modello avrà terminato, verranno mostrate le sue prestazioni selezionando la sua versione.

Dopo tale processo, possiamo eseguire alcuni test (come in Figura 4.10) utilizzando il rilevatore.

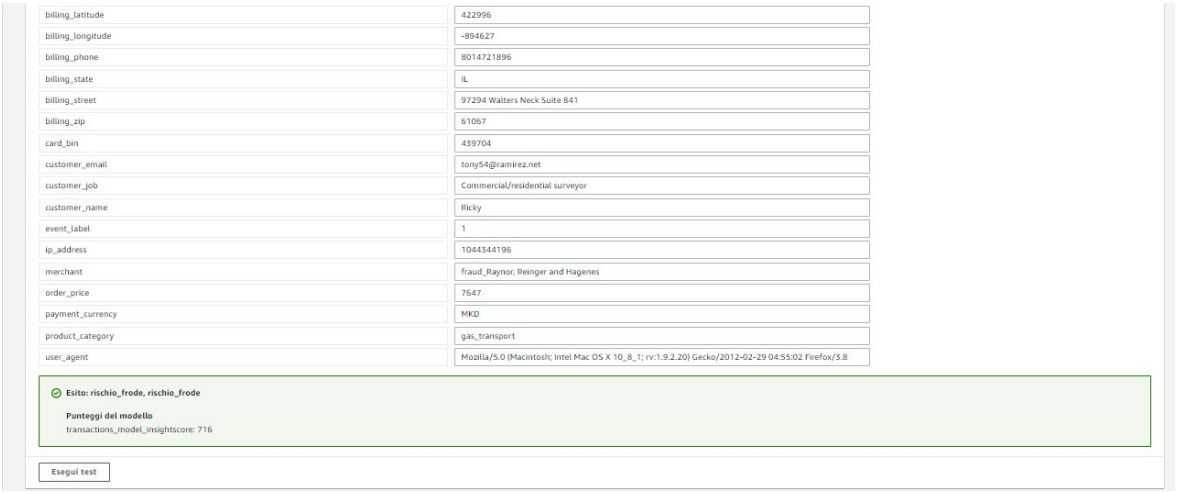


Figura 4.10: Test effettuato sul set di dati usando il rilevatore di frode creato

Alla fine del test, il modello otterrà come risultato un punteggio che corrisponderà alla veridicità dei dati inseriti durante quest'ultima fase e al risultato delle regole configurate, ovvero *rischio_frode* oppure *no_frode*.

CAPITOLO 5

Esperienze sull'Information Retrieval

In questo capitolo verrà illustrato il funzionamento dell'Information Retrieval e come Amazon Kendra, il servizio di ricerca intelligente di AWS, mette in atto questa tecnologia. Successivamente, studieremo come l'NLP ed altre tecnologie avanzate influiscono sulla ricerca delle informazioni e quali sono i benefici che essi offrono. Infine, analizzeremo il futuro della tecnologia di Kendra.

5.1 Introduzione all'Information Retrieval

L'*Information Retrieval (IR)*, ovvero il recupero delle informazioni, si occupa di organizzare, elaborare e accedere a enormi quantità di risorse per estrarre informazioni rilevanti in risposta a una query dell'utente.

Nel corso degli anni, il campo del recupero delle informazioni ha subito notevoli evoluzioni. Inizialmente, il focus era sul recupero di sole risorse testuali, ma oggi si è ampliato per includere articoli bibliografici, contenuti multimediali (come testi, audio, immagini e video), e dati archiviati in database. La capacità di gestire e recuperare informazioni da contenuti multimediali ha portato allo sviluppo di nuovi strumenti e tecniche specializzate.

A differenza dei sistemi di *Data Retrieval*, che gestiscono dati con una struttura ben definita e restituiscono risultati precisi, i sistemi di IR trattano testi scritti in linguaggio naturale, spesso non ben strutturati e semanticamente ambigui. Pertanto, l'IR può restituire risultati che non sono sempre esatti o perfettamente pertinenti, ma che possono, comunque, essere utili per l'utente.

5.1.1 Tecniche di Information Retrieval

La tecnica di indicizzazione più utilizzata nell'IR è quella sull'*indice invertito (inverted index)*. Essa consiste nella memorizzazione dell'elenco dei termini contenuti nei documenti della collezione; per ogni termine viene mantenuta una lista dei documenti nei quali esso compare. Questa tecnica è valida solo per le query semplici.

Il processo di indicizzazione inversa si basa su diverse fasi. La fase principale è quella di *tokenizzazione*, in cui il tokenizer trasforma un elemento di testo in un elenco di token; questi ultimi vengono, a loro volta, rielaborati attraveso le seguenti tecniche:

- *eliminazione delle stopword*: questa tecnica elimina gli articoli, le congiunzioni, i verbi frequenti e le particelle pronominali;
- *de-hyphenation*: essa divide le parole contenenti un trattino (ad esempio, nord-est);

- *stemming*: essa si occupa della riduzione dei termini alla loro "radice", ovvero coniuga tutti i verbi all'infinito e rimuove i suffissi e i prefissi;
- *thesauri*: questa tecnica gestisce i sinonimi, gli omonimi e gli iperonimi tramite delle classi di equivalenza predefinite;
- *lemmatizzazione*: essa è simile allo stemming, ma si basa sull'analisi morfologica ed identifica la forma base o il lemma di una parola.

L'applicazione di queste tecniche, però, porta a gestire alcune eccezioni che si possono verificare, come, ad esempio:

- trattini che sono parti integranti di una parola (ad esempio, B-49);
- parole che, se scritte in maiuscolo, assumono un diverso significato (ad esempio, MIT vs la parola tedesca mit);
- punteggiatura che è parte integrante di una parola (ad esempio, 510 D.C.).

5.1.2 Modelli di IR

Formalmente, un modello di Information Retrieval è una quadrupla di elementi indicati con *(D, Q, F, R)*, dove:

- *D* è un insieme di viste logiche dei documenti della collezione;
- *Q* è un insieme di viste logiche (dette query) dei bisogni informativi dell'utente;
- *F* è un sistema per modellare i documenti, le query e le relazioni tra loro;
- *R(qi, dj)* è una funzione di ranking che associa un numero reale non negativo ad una query *qj* e un documento *dj*, definendo un ordinamento tra i documenti con riferimento alla query *qj*.

I modelli principali di IR sono il *modello booleano* e il *modello vettoriale*.

Il modello booleano è il più semplice e si basa sulla teoria degli insiemi e sull'algebra booleana, utilizzando operatori come *AND*, *OR* e *NOT* per definire i criteri di ricerca. Esso recupera solo i documenti che soddisfano esattamente le condizioni specificate e li differenzia in modo binario, ovvero in base al fatto se un documento è rilevante o non rilevante.

Nel modello vettoriale, invece, ogni documento e ogni query sono rappresentati come vettori in uno spazio multidimensionale, in cui ogni dimensione corrisponde ad un termine specifico, con un peso rappresentato da un numero reale positivo. La somiglianza tra documenti e query viene calcolata in base al coseno dell'angolo tra i rispettivi vettori. La ricerca si svolge calcolando il grado di similarità tra il vettore della query e i vettori dei documenti; i documenti con il più alto grado di similarità alla query hanno maggiori probabilità di essere rilevanti per l'utente.

Nelle Figure 5.1 e 5.2 viene illustrato il funzionamento del modello vettoriale.

5.1.3 Recupero delle informazioni tramite Big Data

I *Big Data* si riferiscono a grandi quantità di dati eterogenei, generati a velocità elevata, che richiedono soluzioni tecnologiche avanzate per la loro gestione. La loro integrazione nell'Information Retrieval (IR) ha un notevole potenziale per aumentare l'efficienza del recupero delle informazioni, in particolare di fronte alla crescente mole di dati provenienti

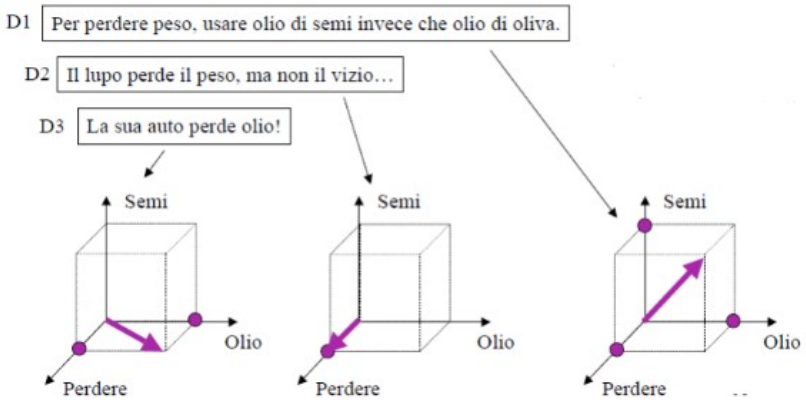


Figura 5.1: Illustrazione dello spazio multidimensionale del modello vettoriale

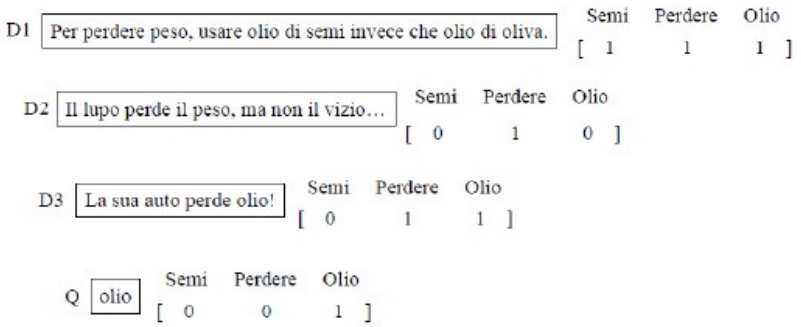


Figura 5.2: Funzionamento del modello vettoriale

dal web e dai social media. L'uso di tecnologie come Apache Hadoop e di piattaforme di analisi distribuite consente ai sistemi di IR di processare rapidamente grandi volumi di dati, migliorando la precisione delle risposte e accelerando il tempo di risposta. Queste soluzioni permettono di gestire dati complessi e favoriscono decisioni più consapevoli.

5.1.4 Parametri di valutazione

Nella valutazione di un sistema di Information Retrieval esistono delle misure standard per valutare la credibilità delle risposte fornite da esso.

I parametri principali sono la Precision, il Recall e l'F1 Score.

La *Precision* valuta la capacità di filtrare documenti non rilevanti in risposta ad una query, mentre il *Recall* indica la percentuale di documenti rilevanti ottenuti come risposta.

Combinando Precision e Recall si ottiene la misura *F1 Score*, che ha il compito di valutare quanto il sistema riesca a bilanciare la qualità (Precision) e la completezza (Recall) dei risultati. Un F1 Score pari a 1 indica un sistema perfetto, mentre un valore vicino a 0 indica scarse prestazioni.

5.1.5 Tecniche di ranking e classificazione

In aggiunta ai parametri di valutazione, abbiamo bisogno di tecniche di ranking per migliorare la ricerca e l'output delle query. Analizziamole di seguito.

- *Ranking probabilistico*: questa tecnica si basa sulla classificazione dei documenti in ordine di probabilità di rilevanza rispetto all'informazione richiesta.

- *Latent Semantic Indexing (LSI)*: nella LSI la ricerca avviene per concetti, ovvero per un insieme di termini correlati detti co-occorrenze o dominio semantico.
- *Relevance Feedback e Query Expansion*: nel Relevance Feedback l'utente seleziona i documenti più rilevanti all'interno di un set aiutando a riformulare la query. Questo passaggio presenterà nuovi risultati all'utente, eventualmente iterando il processo. Nella Query Expansion, invece, si aggiungono ulteriori termini oltre quelli iniziali, con l'obiettivo di migliorare la qualità della ricerca.

5.2 Ruolo dell'elaborazione del linguaggio naturale nell'Information Retrieval (IR)

L'elaborazione del linguaggio naturale (Natural Language Processing, NLP) è un campo interdisciplinare dell'informatica, dell'Intelligenza Artificiale e della linguistica computazionale che si occupa dell'interazione tra i computer e il linguaggio umano.

Infatti, il suo obiettivo è quello di sviluppare sistemi capaci di comprendere, interpretare, generare e manipolare il linguaggio naturale (ossia il linguaggio utilizzato dalle persone nelle comunicazioni quotidiane).

Le tecniche di NLP consentono ai computer di svolgere compiti come la comprensione del testo, il riconoscimento della voce, l'analisi sintattica e semantica, la traduzione automatica e la sintesi vocale. Questi strumenti giocano un ruolo particolare nell'IR poichè contribuiscono a migliorare la comprensione delle query, l'interpretazione dei documenti e la personalizzazione dei risultati. In dettaglio, essi permettono ai motori di ricerca di comprendere le query complesse, le domande esplicite e le frasi interrogative ed aiutano i sistemi a riconoscere l'intento dell'utente dietro le parole digitate.

5.2.1 Benefici dell'NLP nei confronti dell'Information Retrieval

Grazie all'NLP, i sistemi di IR sono migliorati nella comprensione del linguaggio umano e nel rispondere, in modo più intelligente e preciso, alle ricerche effettuate dagli utenti. Vediamo in dettaglio i benefici di questo importante strumento e come esso influenza l'Information Retrieval.

Analisi per significato

Prima dell'introduzione dell'NLP, i sistemi di recupero di informazioni si basavano esclusivamente sulla ricerca delle parole esatte inserite dall'utente. Tuttavia, questo approccio risultava limitato, poichè spesso i documenti rilevanti non venivano trovati se utilizzavano sinonimi o espressioni diverse da quelle digitate. Grazie all'NLP, invece, i motori di ricerca riescono a capire il significato di una query e a cercare anche concetti correlati, migliorando, così, la precisione dei risultati.

Ricerca semantica e Latent Semantic Analysis (LSA)

Oltre alla corrispondenza esatta delle parole, l'elaborazione del linguaggio naturale ha permesso lo sviluppo della *ricerca semantica*, che si basa sulla comprensione dei significati e delle relazioni tra le parole, piuttosto che su semplici corrispondenze letterali.

Una tecnica comune per implementare la ricerca semantica è la *Latent Semantic Analysis (LSA)*, che analizza le relazioni tra un insieme di documenti e i termini in essi contenuti, identificando dei concetti nascosti. Questo approccio consente ai motori di ricerca di restituire documenti pertinenti, anche se non contengono le parole esatte della query.

Per esempio, una query su "energia sostenibile" potrebbe restituire documenti che trattano di "fonti rinnovabili", grazie alla capacità dell'LSA di cogliere le relazioni semantiche tra i termini.

Confrontando l'LSI, citato in precedenza, e l'LSA, possiamo dedurre che sono, di fatto, la stessa cosa. L'unica differenza è che L'LSI viene usato nel recupero delle informazioni, quindi nell'Information Retrieval, mentre l'LSA è più utilizzata nelle scienze cognitive e nella linguistica per modellare e comprendere le relazioni semantiche tra le parole e i concetti che rappresentano. In sintesi, l'LSI può essere visto come un'applicazione specifica dell'LSA in ambito IR.

Espansione della query

L'espansione della query è un'altra applicazione dell'NLP, che permette al sistema di modificare o arricchire la domanda per ottenere risultati migliori. Ad esempio, cercando "auto elettrica", il sistema potrebbe automaticamente trovare e considerare validi anche sinonimi come "veicolo elettrico" o altre espressioni correlate, aumentandone la precisione. Inoltre, utilizzando tecniche come il feedback di rilevanza, il sistema può adattare la ricerca in base ai risultati che inizialmente sembrano più utili all'utente, migliorando ulteriormente le risposte.

Organizzare e classificare i documenti

Inoltre, l'NLP è utile per organizzare i documenti in modo intelligente. Ad esempio, applicando tecniche come il Named Entity Recognition (NER), è possibile di riconoscere automaticamente i nomi di persona, i luoghi o le organizzazioni in un testo, migliorando l'indicizzazione dei documenti. In pratica, il sistema può capire che un documento parla di "Mark Zuckerberg" o di "Meta" ed usarlo per recuperare più facilmente informazioni specifiche.

Migliorare il ranking dei risultati

Uno dei principali obiettivi dell'IR è garantire che i risultati più rilevanti siano mostrati per primi. L'NLP contribuisce a questo miglioramento utilizzando modelli avanzati per analizzare il contesto delle parole nei documenti e per determinare la loro rilevanza rispetto alla query. Ciò significa che il sistema non solo trova i documenti corretti, ma li ordina in base alla loro rilevanza per l'utente, garantendo una migliore esperienza di ricerca.

5.2.2 Approcci moderni all'Information Retrieval basati sull'Intelligenza Artificiale

L'evoluzione del mondo digitale e la conseguente scoperta dell'Intelligenza Artificiale e dei suoi discendenti ha portato tutti i sistemi digitali alla necessità di cambiamento. L'IA è diventato il "cuore" dell'Information Retrieval offrendo un focus sulle tecnologie più recenti, come i modelli basati sulle reti neurali e il loro impatto sul recupero delle informazioni.

Modelli di Linguaggio Avanzati: trasformatori, BERT e GPT

Una delle innovazioni più significative nel campo dell'NLP è stata l'introduzione dei modelli basati sui trasformatori, quali BERT (Bidirectional Encoder Representations from Transformers) e GPT (Generative Pre-trained Transformer). Questi modelli hanno trasformato il modo in cui le macchine comprendono e processano il linguaggio naturale, permettendo una rappresentazione bidirezionale del contesto in cui le parole appaiono; in altre parole,

esse considerano sia ciò che precede che ciò che segue una parola così da determinarne il significato in modo più accurato.

In particolare, BERT è un modello di apprendimento automatico basato su trasformatori ed utilizzato nell'elaborazione del linguaggio naturale (NLP).

BERT è basato sull'architettura dei Transformer, un modello di rete neurale introdotto da Google nel 2017. Sebbene il Transformer utilizzi sia una rete di encoder che una di decoder, BERT impiega esclusivamente la rete di encoder multilivello, poiché è un modello di pre-addestramento, con l'obiettivo di apprendere delle rappresentazioni linguistiche a partire dai dati di input. Il Transformer si distingue per il modulo di multi-head attention, che ha trovato applicazioni di successo sia nel campo della Computer Vision che in quello dell'elaborazione del linguaggio naturale.

In questo contesto, Google introduce due versioni del modello BERT, ovvero:

- *BERT-base*: esso è composto da 12 encoder e 110 milioni di parametri (12 head di auto-attention bidirezionali);
- *BERT-large*: esso è composto da 24 encoder e 340 milioni di parametri (16 head).

In sintesi, BERT associa ad ogni parola una rappresentazione numerica, nota come *embedding*. Ogni livello della rete calcola l'attenzione sugli embedding del livello precedente, producendo nuovi embedding intermedi. Questi mantengono la stessa dimensione di quelli originari.

In Figura 5.3 è raffigurato un esempio di architettura BERT che include le due versioni del modello.

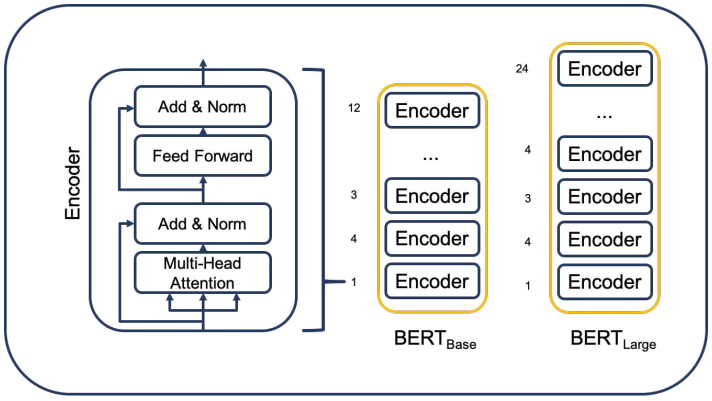


Figura 5.3: Architettura BERT

GPT, d'altra parte, ha dimostrato come i modelli generativi possano essere utilizzati non solo per comprendere il linguaggio, ma anche per generare risposte a query complesse. Ciò apre la porta ad una nuova generazione di sistemi di recupero che possono restituire non solo documenti, ma anche risposte elaborate in linguaggio naturale.

Un altro pilastro dell'NLP nell'IR sono le rappresentazioni contestuali delle parole, come quelle fornite dai modelli di embedding, ovvero Word2Vec, GloVe ed ELMo. Questi modelli usano il modello di Information Retrieval di tipo vettoriale.

Ad esempio, una query contenente la parola "automobile" restituirà anche documenti che parlano di "auto" o "veicolo", grazie alla somiglianza semantica tra questi termini. I modelli come ELMo, invece, sono ancora più avanzati, poichè producono degli embedding che variano a seconda del contesto in cui le parole appaiono. Ciò è particolarmente utile per risolvere le ambiguità semantiche in cui una parola può avere più significati a seconda della frase.

L'integrazione delle tecniche di NLP nell'Information Retrieval ha portato a miglioramenti significativi nella capacità dei sistemi di comprendere il linguaggio umano, gestire delle query complesse e restituire i risultati pertinenti. Infatti, dai modelli avanzati, come BERT e GPT, alle tecniche di analisi semantica e riconoscimento delle entità, l'NLP ha ampliato enormemente le possibilità di migliorare la precisione, la pertinenza e la personalizzazione delle ricerche. Questi sviluppi continuano ad evolvere, offrendo delle prospettive entusiasmanti per il futuro dell'Information Retrieval.

5.3 Amazon Kendra

Amazon Kendra è un servizio di ricerca intelligente offerto dalla piattaforma AWS, basato sul Machine Learning. Esso permette agli utenti di eseguire delle ricerche in diversi repository di contenuti configurati attraverso i connettori integrati al suo interno. Amazon Kendra è adatto per la ricerca nei siti Web e nelle applicazioni permettendo, così, a dipendenti e ai clienti di trovare i contenuti che cercano, anche quando sono sparsi in più posizioni e repository di contenuti all'interno dell'organizzazione, nel più breve tempo possibile. Inoltre, questo servizio utilizza l'elaborazione del linguaggio naturale e gli algoritmi avanzati di apprendimento automatico per restituire delle risposte specifiche a domande di ricerca sui propri dati.

5.3.1 Come funziona Amazon Kendra

Amazon Kendra indicizza i documenti direttamente o da un repository di documenti di terze parti e fornisce in modo intelligente informazioni rilevanti agli utenti. Il servizio usa le sue capacità di comprensione semantica e contestuale per decidere se un documento è pertinente ad una query di ricerca o meno. Di conseguenza, Amazon Kendra restituisce risposte specifiche alle domande, offrendo agli utenti un'esperienza che è vicina all'interazione con un esperto umano.

In Amazon Kendra è possibile porre diverse tipologie di domande:

- *Domande factoid*: le domande factoid forniscono risposte basate sui fatti che possono essere restituite come una singola parola o frase. La risposta viene recuperata da una FAQ o dai documenti indicizzati.
- *Domande descrittive*: queste sono domande in cui la risposta potrebbe essere una frase, un brano o un intero documento.
- *Domande su parole chiave e linguaggio naturale*: queste sono domande che includono contenuti complessi e colloquiali il cui significato potrebbe non essere chiaro. Ad esempio, considerando la frase *discorso di apertura*, oppure, nel caso in cui Amazon Kendra incontra una parola come "indirizzo", che ha più significati contestuali, esso deduce correttamente il significato in base alla query di ricerca e restituisce, di conseguenza, informazioni pertinenti.

Questo servizio di ricerca permette di scegliere una fonte di dati configurando diversi connettori inclusi in AWS, ovvero:

- *Amazon S3*: attraverso il bucket Amazon S3, Kendra può gestire e cercare tra i file memorizzati nel servizio di storage di AWS.
- *Microsoft SharePoint*: esso consente di connettere Amazon Kendra ai repository di SharePoint, inclusi SharePoint Online e SharePoint Server, contenenti documenti, pagine e liste.

- *Amazon RDS (Amazon Relational Database) e Amazon Aurora*: essi connettono Kendra a database relazionali, ospitati sulle loro piattaforme, per indicizzare e cercare dei dati strutturati, come, ad esempio, le tabelle e i record di database relazionali.
- *Salesforce*: esso include gli oggetti, i record e i documenti conservati nella piattaforma di CRM, quali opportunità, contatti e casi.
- *JIRA*: esso consente di indicizzare i ticket e i problemi gestiti nel cloud JIRA, migliorando l'accesso alle informazioni sui progetti e sulle attività.
- *Google Drive*: esso contiene i file e le cartelle condivise o archiviate nel cloud di Google.
- *Web Crawler*: è un connettore in grado di navigare automaticamente e raccogliere contenuti da siti web pubblici specificati.
- *Connettori personalizzati*: Amazon Kendra consente la creazione di connettori personalizzati per integrare fonti di dati specifiche che non sono coperte dai connettori predefiniti.

Come ultimo step, si passa alla sincronizzazione e all'indicizzazione definitiva dei documenti presenti al loro interno, così da avviare la ricerca.

Nella prossima sezione vedremo come questa tecnologia si differenzi nettamente dai metodi di ricerca tradizionali e come offra significativi vantaggi nell'estrazione delle informazioni.

5.3.2 IA generativa vs ricerca tradizionale

Amazon Kendra è una soluzione basata sull'apprendimento automatico che offre una serie di vantaggi rispetto alle tradizionali tecnologie di ricerca per parole chiave, soprattutto quando si tratta di analizzare dei dati non strutturati, come i PDF, i documenti Word e le pagine HTML.

In particolare, esso fornisce risposte più accurate alle query di ricerca, migliorando continuamente grazie all'aggiornamento dei suoi modelli; infine, esso è semplice da implementare. Questa tecnologia offre risultati più pertinenti ed accurati anche per le domande complesse ed è progettata per gestire grandi quantità di dati, inclusi i documenti aziendali, le FAQ, i manuali e altro ancora.

A differenza dei metodi di ricerca tradizionali, che si basano principalmente su parole chiave e richiedono all'utente di formulare query specifiche, Amazon Kendra è in grado di estrarre risposte dirette dai documenti, piuttosto che restituire semplicemente una lista di documenti rilevanti. Al contrario, i sistemi di ricerca tradizionali possono fornire risultati meno pertinenti se le parole chiave non sono scelte con cura, poiché mancano di una comprensione profonda del contesto o dell'intento della ricerca.

5.3.3 Applicazione di Amazon Kendra

Lo step principale per la creazione di un modello in Amazon Kendra consiste nel creare un'indice e collegare una fonte da cui prendere i dati. In questo caso abbiamo selezionato il bucket S3 di AWS chiamato *documentilegislativi*, contenente, appunto, una serie di documenti legislativi (Figura 5.4) in formato `.pdf`.

In seguito, per un'ulteriore test, abbiamo sperimentato anche il connettore WebCrawler, che preleva informazioni direttamente dal web, in particolare da uno o più URL della fonte di dati che abbiamo inserito, ovvero il sito della Gazzetta Ufficiale dello Stato, da cui prendere gli stessi documenti, ma in formato `.HTML`.

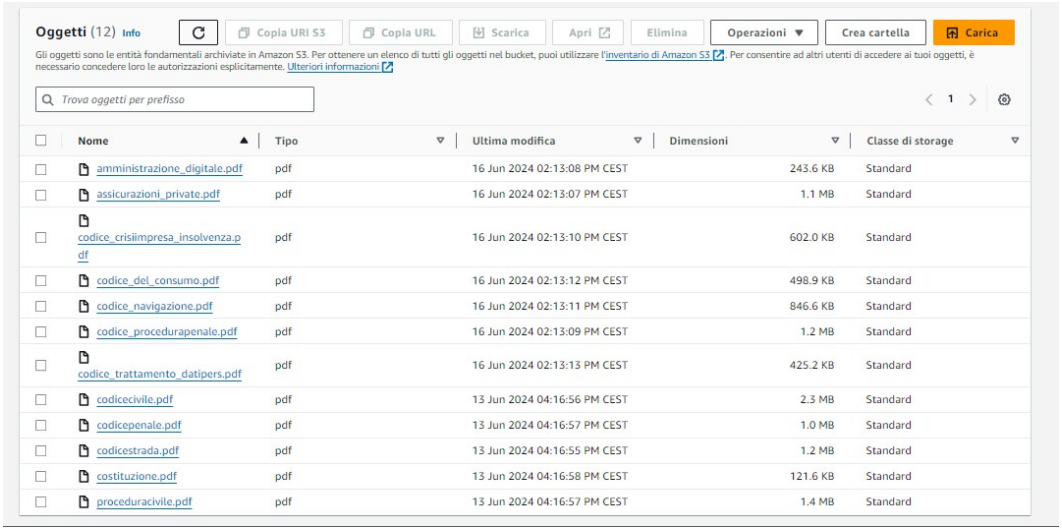


Figura 5.4: Documenti presenti nel bucket S3 da indicizzare

Amazon Kendra, successivamente, provvederà ad indicizzare questi documenti secondo una serie di campi indici predefiniti, come l'autore, la categoria, la data di creazione del documento, le informazioni sulla struttura del testo (titolo e corpo) e la lingua. È possibile inserire altri campi per migliorare questa fase; infatti, nel nostro modello, abbiamo inserito un suggerimento sulla struttura dei nostri file, come, ad esempio, le variabili *num_articolo* e *sezione* riferite alle leggi costituzionali.

In questo contesto, l'approccio analizzato consente, ad esempio, ad un avvocato o ad un funzionario pubblico di trovare rapidamente articoli di legge specifici all'interno di un vasto archivio di documenti legislativi. Anzichè dover sfogliare manualmente centinaia di file, l'utente può formulare una semplice query, come "articolo 5 della Costituzione", e Kendra restituirà immediatamente la porzione esatta del testo in cui è citato, riducendo notevolmente i tempi di ricerca.

Il servizio AWS, una volta indicizzati tutti i documenti, permetterà di creare un'*experience*, ovvero la nostra pagina di ricerca, che avrà la struttura di un motore di ricerca Google (Figura 5.5).

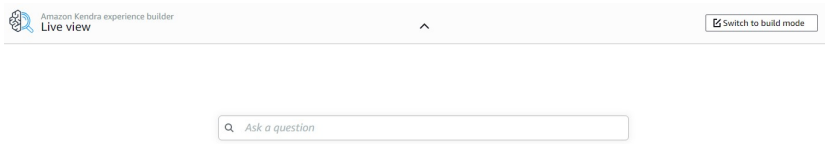


Figura 5.5: Experience di Amazon Kendra

Essa può essere personalizzata, se lo si desidera, inserendo le FAQ o le parole chiavi per aiutare l'utente a capire quali argomenti ricercare e se sta consultando la documentazione corretta.

Per accedere a questa schermata, Amazon Kendra richiede delle credenziali di accesso (Figura 5.6), configurate in precedenza, composte da un nome utente e una password.

Figura 5.6: Schermata di accesso all'experience

A questo punto, una volta effettuato l'accesso, possiamo testare la nostra experience per effettuare le ricerce sui documenti legislativi inseriti. Ad esempio, cercando la parola "Repubblica", come in Figura 5.7, Amazon Kendra restituirà il nome del file contenente la parola inserita, in questo caso `costituzione.pdf` e la parte di testo associata.

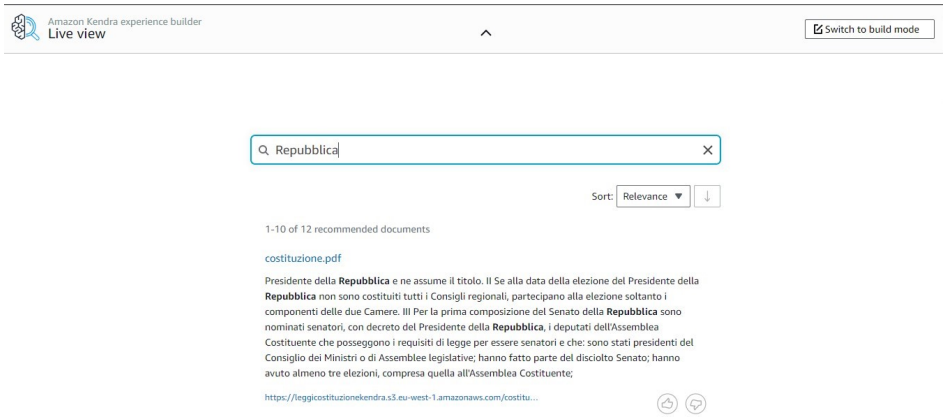


Figura 5.7: Ricerca con Amazon Kendra

Come abbiamo potuto osservare, Amazon Kendra è uno strumento molto potente per ricercare velocemente le parole presenti in una grande quantità di dati. La sua implementazione nelle aziende o in contesti specifici, come quello legislativo, permette di risparmiare tempo, migliorare l'efficienza e garantire che le informazioni rilevanti siano facilmente accessibili.

Guardando al futuro, le potenzialità di Amazon Kendra hanno molta probabilità di espandersi ulteriormente, con sviluppi che potrebbero ampliare significativamente le sue applicazioni e migliorare le sue capacità. Questi progressi potrebbero portare ad una maggiore precisione nelle ricerche e ad una comprensione più profonda delle query complesse, permettendo a Kendra di fornire risposte ancora più mirate e contestualizzate. Questo grazie anche all'integrazione di tecnologie come l'Intelligenza Artficiale avanzata, l'apprendimento automatico e l'elaborazione del linguaggio naturale.

5.3.4 Prospettive future

In definitiva, l'integrazione dell'NLP nell'Information Retrieval rappresenta un passo fondamentale per rispondere alla complessità crescente delle query utente e dell'informazione disponibile. Con tecniche sempre più avanzate, i sistemi di IR diventano più intelligenti, perso-

nalizzati e capaci di offrire risultati pertinenti in tempi rapidi, migliorando significativamente l'esperienza di ricerca.

Amazon Kendra ha già rivoluzionato il modo in cui le organizzazioni gestiscono e accedono alle informazioni. Tuttavia, il suo futuro è particolarmente promettente, con diversi sviluppi in vista che potrebbero ampliare le sue capacità e le sue applicazioni. Alcune delle aree in cui Amazon Kendra potrebbe evolversi nei prossimi anni includono:

- *Integrazione con AI generativa*: l'adozione di modelli di linguaggio avanzati, come GPT-4 e le sue versioni future, potrebbe migliorare significativamente la capacità di Kendra di generare risposte contestualizzate e più dettagliate. Questo tipo di AI potrebbe permettere a Kendra non solo di restituire risposte basate sui documenti esistenti, ma anche di generare spiegazioni o riassunti più complessi, offrendo un'interazione più naturale e intuitiva con gli utenti.

- *Miglioramenti nella personalizzazione*: le tecnologie di personalizzazione stanno diventando sempre più sofisticate, e Amazon Kendra non è da meno. Le recenti innovazioni nel campo dell'apprendimento automatico hanno reso possibile una personalizzazione più fine delle esperienze di ricerca. Kendra potrebbe sfruttare questi sviluppi per adattare i risultati della ricerca in base ai comportamenti e alle preferenze degli utenti, offrendo un'esperienza di ricerca altamente personalizzata.

- *Espansione dell'integrazione con altre tecnologie*: Amazon Kendra si sta già integrando con numerosi servizi AWS e altre piattaforme. In futuro, si prevede che questa integrazione si estenderà ulteriormente. Per esempio, potrebbero essere sviluppati connettori più avanzati per integrare Kendra con nuovi tipi di fonti di dati e applicazioni di terze parti. Questa espansione potrebbe rendere Kendra uno strumento ancora più versatile, capace di interagire con una vasta gamma di sistemi aziendali e di piattaforme di comunicazione.

- *Innovazioni nella sicurezza dei dati*: con l'aumento delle preoccupazioni riguardanti la sicurezza dei dati, l'integrazione di tecnologie come la blockchain potrebbe diventare un'opzione per garantire l'integrità e la protezione delle informazioni gestite dal servizio AWS. Sebbene questa integrazione sia ancora nelle fasi iniziali di esplorazione, potrebbe offrire soluzioni avanzate per la verifica e la protezione dei dati.

- *Scalabilità e prestazioni*: Amazon Kendra gestisce ed indicizza enormi volumi di dati. Con l'espansione continua delle fonti di dati e l'aumento delle richieste, la scalabilità assume un ruolo molto importante. Per questo motivo, le future versioni di Kendra potrebbero includere miglioramenti nella gestione della scalabilità, per garantire che il servizio possa continuare ad offrire prestazioni elevate anche con set di dati sempre più grandi e complessi.

In sintesi, Amazon Kendra rappresenta una soluzione innovativa nel campo dell'Information Retrieval, con prospettive future che promettono di espandere ulteriormente le sue capacità e applicazioni. Kendra è, quindi, destinato a rimanere un punto di riferimento nell'evoluzione della gestione dei dati e nella ricerca intelligente.

CAPITOLO 6

Discussione

In questo capitolo discuteremo in merito alle esperienze effettuate ed analizzate nei capitoli precedenti per ciò che riguarda il riconoscimento di immagini, il rilevamento delle frodi e il recupero di informazioni. Per ciascuna di tali esperienze trarremo alcune considerazioni.

6.1 Riconoscimento di immagini con Amazon Rekogntion: esperienza e considerazioni

Il riconoscimento delle immagini è uno degli ambiti più avanzati dell'Intelligenza Artificiale. Amazon Rekognition, uno dei servizi di Amazion Web Service, grazie alle sue funzionalità di Deep Learning, consente l'identificazione di oggetti, persone e scene in modo preciso ed efficiente.

Durante l'esperienza pratica, abbiamo testato Rekognition tramite la console della piattaforma, sfruttando sia le demo preaddestrate che la funzionalità *Custom Labels*, che consente di creare modelli specifici per riconoscere categorie particolari di immagini. Le demo, invece, hanno evidenziato l'abilità del servizio nel riconoscimento facciale e nella moderazione dei contenuti, dimostrando la sua utilità per settori come la sicurezza pubblica e l'analisi dei contenuti sui social media.

L'aspetto più interessante è stato l'uso di *Custom Labels*, che permette di addestrare modelli personalizzati con set di dati specifici, nel nostro caso abbiamo preso in considerazione varie specie di orchidee. La qualità e la risoluzione delle immagini hanno mostrato un impatto significativo sulle predizioni.

Un altro elemento rilevante emerso durante l'esperienza con Amazon Rekognition è la semplicità d'uso della console AWS. Nonostante la complessità della tecnologia sottostante, l'interfaccia è intuitiva e permette agli utenti di eseguire operazioni avanzate come il caricamento del set di dati, l'etichettatura delle immagini e l'addestramento di modelli, senza richiedere competenze di programmazione approfondite. L'unico requisito per ottenere un'analisi molto accurata è quello di avere un set di immagini sufficientemente ampio e ben strutturato.

Inoltre, l'integrazione con altri servizi AWS, come Amazon S3 per la gestione dei dati, facilita il salvataggio e la reperibilità delle immagini. Ciò rende Amazon Rekognition non solo uno strumento potente per il riconoscimento delle immagini, ma anche uno strumento altamente scalabile ed integrabile in diversi flussi di lavoro aziendali.

Un ultimo aspetto interessante da considerare è il continuo miglioramento del servizio grazie a nuove tecniche di Machine Learning, che consentono ai modelli di adattarsi e migliorare nel tempo, ottimizzando la precisione delle predizioni e ampliando le possibilità di utilizzo in settori in rapida evoluzione, come la videosorveglianza e l'analisi dei media.

6.2 Rilevamento delle frodi con Amazon Fraud Detector: esperienza e considerazioni

Il rilevamento delle frodi rappresenta una sfida cruciale per settori come quello finanziario e dell'e-commerce, dove le transazioni sospette sono in costante aumento.

Amazon Fraud Detector sfrutta i modelli di Machine Learning per analizzare i dati in tempo reale e prevenire attività fraudolente, riducendo, allo stesso tempo, i falsi positivi e migliorando l'efficienza operativa.

Durante l'esperienza pratica effettuata con Fraud Detector, abbiamo caricato un set di dati storici di transazioni e configurato regole personalizzate per rilevare operazioni anomale. La console di AWS, con la sua interfaccia user-friendly, ha reso semplice definire le variabili ed impostare etichette per identificare eventi sospetti. Inoltre, il processo di configurazione dei modelli è stato rapido mostrando, successivamente, un'analisi dettagliata delle performance sulla dashboard, consentendo il monitoraggio della precisione e dell'accuratezza del modello di rilevamento delle frodi, attraverso la metrica AUC.

I risultati ottenuti hanno dimostrato un'alta affidabilità del sistema, riducendo significativamente i tempi di rilevamento e migliorando la capacità di identificare frodi in tempo reale.

L'aspetto più interessante di questa analisi è stato, sicuramente, la possibilità di riaddestrare i modelli man mano che vengono aggiunti nuovi dati, migliorando progressivamente le performance del sistema.

Tuttavia, un fattore critico emerso riguarda la necessità di disporre di set di dati transazionali ampi, ben strutturati e ben definiti, per garantire un'efficacia ottimale.

In conclusione, Amazon Fraud Detector si è rivelato uno strumento potente e versatile per il rilevamento delle frodi, con la capacità di adattarsi ai cambiamenti e di evolvere costantemente, migliorando la sicurezza e l'affidabilità in diversi settori.

6.3 Rilevamento di testi con Amazon Kendra: esperienza e considerazioni

Amazon Kendra rappresenta un avanzamento significativo nel campo dell'Information Retrieval (IR), grazie alla sua capacità di comprendere le query in linguaggio naturale utilizzando tecniche di Natural Language Processing (NLP). Questo rende Kendra uno strumento essenziale per migliorare la gestione della conoscenza aziendale e per ottimizzare il recupero di informazioni da grandi quantità di dati non strutturati.

Indicizzando i documenti presenti nel bucket Amazon S3, il sistema si è dimostrato altamente efficiente nel fornire risposte pertinenti anche a domande complesse, riducendo i tempi di ricerca e migliorando la precisione.

Una delle caratteristiche più rilevanti di Kendra è la capacità di connettersi a fonti di dati eterogenee e di restituire dei risultati basati non solo su parole chiave, ma anche sull'intento dell'utente, migliorando notevolmente l'efficacia delle ricerche.

Anche in questo caso, l'uso della console AWS si è dimostrato molto vantaggioso e ha reso semplice la configurazione delle diverse fonti di dati utilizzate. Le funzionalità di monitoraggio e reporting della console, invece, hanno permesso di analizzare le performance

del sistema con l'uso dell'experience di ricerca e di ottimizzare i risultati ottenuti con facilità. Non sono state riscontrate difficoltà ed ostacoli nell'esecuzione del recupero di informazioni. Ciò dimostra quanto sia veramente potente ed efficace effettuare queste operazioni con il servizio di Amazon Web Service, rendendo tale servizio ideale per organizzazioni che gestiscono grandi volumi di dati. Tuttavia, l'aggiornamento delle connessioni rimane un'area in cui il sistema potrebbe essere ulteriormente ottimizzato, magari con l'uso di ulteriori tecnologie all'avanguardia.

Conclusioni e sviluppi futuri

In questa tesi si è partiti dalla definizione del concetto di Intelligenza Artificiale, analizzando in dettaglio le tecnologie che derivano da essa e le varie tipologie di IA, ovvero l'IA debole e l'IA forte, per poi esplorare una serie di esperimenti mentali, atti a categorizzare una macchina come intelligente o meno, come il test di Turing e quello della stanza cinese. Per capire meglio il funzionamento di questa enorme risorsa, si è passati allo studio dei vari campi in cui essa può essere applicata, dalla medicina alle tecnologie IT.

L'analisi svolta, inoltre, ci ha permesso di esplorare l'applicazione dell'Intelligenza Artificiale in contesti di grande rilevanza, quali il riconoscimento di immagini, il rilevamento delle frodi e il recupero delle informazioni, fornendo una visione approfondita delle potenzialità offerte da strumenti innovativi quali Amazon Rekognition, Amazon Fraud Detector e Amazon Kendra. Per ciascun contesto, ci siamo soffermati su un'introduzione esplicativa dei servizi utilizzati, per poi testare le funzionalità delle tecnologie di Machine Learning offerte da AWS.

Il lavoro che abbiamo sviluppato in questa tesi si conclude, infine, con una discussione sulle esperienze di utilizzo dell'Intelligenza Artificiale da noi compiute. Con il tempo, l'IA si è affermata sempre di più nelle nostre vite rivoluzionando numerosi settori e portando, di conseguenza, innovazioni che influenzeranno soprattutto la vita quotidiana e il mondo del lavoro. Ciò porterà alla definizione di vantaggi, ma anche di svantaggi.

L'Intelligenza Artificiale ha compiuto significativi progressi dai suoi esordi negli anni Novanta, quando si concentrava sulla correlazione di dati attraverso regole predefinite. Con l'aumento dei dati disponibili e la crescente domanda di risposte rapide, l'IA è evoluta verso il Machine Learning, passando dalla programmazione all'addestramento. Questo approccio ha consentito ai sistemi di apprendere da dati di input e da risposte note, generando regole applicabili a situazioni simili.

Ulteriori sviluppi hanno portato a tecnologie come il Deep Learning, che hanno ampliato le applicazioni dell'IA in vari ambiti, dalla robotica alla diagnostica medica, passando per le traduzioni automatiche e il riconoscimento delle immagini. Tuttavia, attualmente l'IA è spesso percepita come una "scatola nera", con poca trasparenza sui processi decisionali interni, il che richiede l'implementazione di meccanismi di trasparenza per costruire fiducia in questa tecnologia.

In ambito economico, si prevede un aumento degli investimenti in IA, passando da circa 2 miliardi di euro nel 2016 a 60 miliardi nel 2025, con la Cina destinata a diventare il leader mondiale nel settore entro il 2030. Per l'Unione Europea, la sfida non è tanto vincere questa corsa, quanto guidare lo sviluppo dell'IA in modo etico e rispettoso dei valori umani.

Il mondo del lavoro subirà cambiamenti radicali; infatti, le mansioni tradizionali saranno sostituite da soluzioni automatizzate, come gli assistenti vocali e i veicoli autonomi. Sebbene l'IA sostituirà lavori a bassa qualificazione, si prevede che creerà nuove opportunità, specialmente in ambito digitale, con una richiesta stimata di oltre 1,7 milioni di posti di lavoro in questo settore entro il 2030. La digitalizzazione richiederà competenze più elevate, ma comporterà anche un rischio di disuguaglianza socio-economica.

I contratti sociali dovranno essere rivisti per affrontare le disuguaglianze emergenti, con possibili soluzioni che includano forme di stipendio base universale e un quadro regolamentare che protegga i diritti dei lavoratori e dei consumatori.

In definitiva, il futuro dell'IA dovrà essere costruito in modo tale da riflettere e supportare l'umanità, rendendo la tecnologia uno strumento di progresso e inclusione.

Bibliografia

C. C. AGGARWAL. *Neural networks and deep learning: A textbook*. Springer, 2018.

B. BAESENS, V. VAN VLASSELAER, and W. VERBEKE. *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. John Wiley & Sons, 2015.

N. BOSTROM. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

T. BRANTS. Natural language processing in information retrieval. *The Clinician*, 2003.

L. CABIBBO. Cloud computing. Dispensa ASW 470, ottobre 2014.

G. G. CHOWDHURY. *Introduction to modern information retrieval*. Facet publishing, 2010.

G. D'ONZA. *La prevenzione delle frodi aziendali: Alle radici della responsabilità sociale*. Franco Angeli Edizioni, 2013.

M. EGMONT-PETERSEN, D. DE RIDDER, and H. HANDELS. Image processing with neural networks—a review. *Pattern recognition*, 35(10):2279–2301, 2002.

S. GEE. *Fraud and Fraud Detection: A Data Analytics Approach*. John Wiley & Sons, 2014.

S. HAYKIN. *Neural Networks and Learning Machines, Third Edition*. Pearson, 2009.

R.A. KAMBAU and Z.A. HASIBUAN. Evolution of information retrieval system: Critical review of multimedia information retrieval system based on content, context, and concept. In *11th International Conference on Information Communication Technology and System (ICTS)*, pages 91–98, 2017.

G. MEINI, F. FORMICHI, M. SARTOR, and A. PARODI. *Corso di sistemi e reti: Virtualizzazione dei sistemi, sicurezza e gestione delle reti e Internet of Things*. Zanichelli, 2022.

M. MITCHELL. Artificial intelligence—a guide for thinking humans. *Genetic Programming and Evolvable Machines*, 2022.

P. NORVIG and S. RUSSEL. *Artificial Intelligence: A Modern Approach Third Edition*. Pearson, 2009.

P. NORVIG and S. RUSSEL. *Artificial Intelligence: A Modern Approach Fourth Edition*. Pearson, 2021.

MINISTERO DELLA SALUTE. I sistemi di intelligenza artificiale come strumento di supporto alla diagnostica, novembre 2021. Sessione LII.

T.W. SINGLETON and A.J. SINGLETON. *Fraud auditing and forensic accounting*, volume 11. John Wiley & Sons, 2010.

Sitografia

- Akamai – http://www.akamai.com/it/glossary/what-is-multifactor-authentication#: :text=multifattore%20(MFA)%3F-,Che%20cos'%C3%A8%20l'autenticazione%20multifattore%20(MFA)%3F,utenti%20accedono%20ai%20servizi%20online
- Alecsandria Digital – http://www.alecsandria.it/articoli/intelligenza-artificiale-e-lotta-alle-frodi/
- Amazon Web Service – https://aws.amazon.com/it/blogs/machine-learning/prevent-account-takeover-at-login-with-the-new-account-takeover-insights-model-in-amazon-fraud-detector/
- Amazon Web Service – https://aws.amazon.com/it/blogs/machine-learning/detect-fraud-in-mobile-oriented-businesses-using-grabdefence-device-intelligence-and-amazon-fraud-detector/
- Amazon Web Service – https://aws.amazon.com/it/rekognition/
- Amazon Web Service – https://docs.aws.amazon.com/it_it/rekognition/latest/dg/how-it-works.html
- Amazon Web Service – https://docs.aws.amazon.com/it_it/whitepapers/latest/aws-overview/types-of-cloud-computing.html
- Amazon Web Service– https://aws.amazon.com/it/blogs/machine-learning/introducing-amazon-kendra/
- Andrea Minini – http://www.andreaminini.com/ai/le-reti-neurali-informatiche
- DEFENSIS – http://www.defensis.it
- Forbes – https://www.forbes.com/advisor/it/business/trend-ai-statistiche/
- Google Cloud – https://cloud.google.com/learn/what-is-cloud-architecture?hl=it
- Google Cloud – https://cloud.google.com/learn/what-is-iaas?hl=it#benefits-of-iaas

- IBM – http://www.ibm.com/it-it/topics/artificial-intelligence
- IBM – http://www.ibm.com/it-it/topics/cloud-computing
- Intelligenza Artificiale Italia – http://www.intelligenzaartificialeitalia.net/post/intelligenza-artificiale-forte-e-debole-le-differenze-tra-ia-forte-e-ia-debole
- Istituto Mario Negri – http://marionegri.it/magazine/intelligenza-artificiale-medicina
- LECS – http://lecs.io/proteggersi-dalla-frode-informatica-strumenti-e-consigli-di-sicurezza/
- LinkedIn – http://www.linkedin.com/pulse/la-tecnologia-blockchain-e-le-frodi-informatiche-un-rapporto-andtf/
- MATLAB & Simulink – http://it.mathworks.com/discovery/image-recognition-matlab.html
- Marcovigorelli – http://www.marcovigorelli.org/
- Oracle Italia – http://www.oracle.com/it/applications/what-is-saas/#link3
- Osservatori Blog - Il Blog sull'Innovazione Digitale – http://blog.osservatori.net/it_it/applicazioni-intelligenza-artificiale
- Osservatori Blog - Il Blog sull'Innovazione Digitale – https://blog.osservatori.net/it_it/cloud-computing-significato-vantaggi
- PTC – http://www.ptc.com/it/blogs/iiot/what-is-iot-security
- Red Hat – http://www.redhat.com/it/topics/cloud-computing/what-is-iaas
- Red Hat – https://www.redhat.com/it/topics/cloud-computing/what-is-paas
- WIKIBOOKS – https://it.wikibooks.org/wiki/Intelligenza_artificiale/Stanza_cinese
- Wikipedia – http://en.wikipedia.org/wiki/Amazon_Rekognition
- Wikipedia – http://it.wikipedia.org/wiki/Cloud_computing
- AI News – http://ainews.it/cose-lintelligent-data-processing/

Ringraziamenti

Un ringraziamento speciale va alla mia famiglia. Mi avete supportata nei momenti di crollo e nei momenti di felicità. Vi chiedo scusa per tutte quelle volte in cui i "sono stata bocciata" superavano i "sono stata promossa"...Lo so, vi ho fatto arrabbiare un pò, ma eccomi qui, finalmente Ingegnere!!! Spero siate soddisfatti di questo mio grande traguardo, realizzato soprattutto grazie a voi.

Un altro importantissimo ringraziamento va al Prof. Domenico Ursino, nonchè relatore della mia tesi. La ringrazio di cuore per la sua estrema disponibilità e simpatia. Grazie per avermi accompagnata durante questo percorso seguendomi costantemente, anche in pieno agosto. La sua professionalità e dedizione sono per me una fonte d'ispirazione.

Ai miei angeli, le mie stelle nel cielo, che da lassù mi osservano e mi sostengono. Spero di avervi resi orgogliosi di me.

Ringrazio di cuore i miei compagni di corso: Alessandra, Alessio, Edoardo, Giansimone, Laura, Luca e Walter per il supporto che mi avete fornito in questi anni. Grazie per le tante risate e il tempo passato insieme; presto recuperermo tutte le cene che non abbiamo fatto in quest'ultimo anno.

A Mariapia, un'amica speciale con cui ho condiviso momenti indimenticabili. Ti penso sempre, anche se ultimamente ci siamo allontanate un pò. Ti prometto che nonostante i vari impegni, recupereremo il tempo perso. In bocca al lupo per il tuo futuro!

Alle mie amiche d'infanzia; nonostante tutto e dopo tutti questi anni siamo ancora insieme. Anche a voi, mando un grande in bocca al lupo per il vostro futuro.

Al Prof. D'Ortona, grazie per esserci sempre e grazie per ricordarmi di vedere sempre il bicchiere mezzo pieno. Sei il prof che tutti vorrebbero avere, mi hai visto crescere nel mio percorso scolastico e mi hai sostenuto in quello universitario. Ora, però, possiamo finalmente definirci "colleghi".

Ai miei coinquilini presenti e passati, Lorenzo, Lisa, Francesco, Francesca e Luca, grazie per aver condiviso con me gioie e dolori e per rendere meno noiosa la vita di tutti i giorni.

A te Lisa, dedico un pensiero particolare, ti ringrazio perchè sei stata la miglior coinquilina di sempre; quando sei andata via ammetto che mi è scappata qualche lacrimuccia... Grazie per essere stata sempre mia complice e per aver condiviso con me i vari scleri, tu sai ahaha.