Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea in Informatica

SVILUPPO DI UN APPLICATIVO WEB PER LA RILEVAZIONE PREVENTIVA DI PRODOTTI ILLEGALI E FRAUDOLENTI NEGLI E-COMMERCE

Relatore:
Chiar.mo Prof.
ANGELO DI IORIO

Correlatore:
GIOVANNI BRUNO

Presentata da:
SIMONE RUGGIERO

I Sessione
Anno Accademico 2023/2024

Abstract

Con l'aumento esponenziale degli acquisti online, il rischio di frodi e truffe è diventato un problema sempre più rilevante per i marketplace che sono costantemente alla ricerca di nuove soluzioni per garantire e mantenere il rapporto fiduciario (trust) con i loro clienti.
Questo lavoro di tesi si propone di analizzare e valutare l'efficacia di un'applicazione web basata su intelligenza artificiale, per il riconoscimento preventivo di prodotti fraudolenti in siti di e-commerce. L'applicativo, sviluppato in occasione di un tirocinio formativo, è diventato un prodotto di mercato, a dimostrazione dell'elevata richiesta oggi presente in questo settore.

L'applicativo permette di analizzare i prodotti prima che questi siano pubblicati nel marketplace, inviandoli ad un modello di intelligenza artificiale per essere valutati e di confermare o rigettare i risultati proposti. Alcuni modelli, come GPT, Claude e Gemini, grazie alla capacità di autoapprendimento, permettono di migliorare le proprie prestazioni nel tempo, apprendendo dai feedback ricevuti. I risultati ottenuti dimostrano come questi modelli siano attualmente la scelta migliore per il riconoscimento di prodotti fraudolenti, con un'accuratezza che può raggiungere, in alcuni casi, perfino valori intorno al 99%.

L'interazione umana con l'applicativo, soprattutto nella sua prima fase di utilizzo, costituisce un elemento fondamentale per garantire la qualità dei risultati ottenuti, ma l'obiettivo finale è quello di rendere l'applicativo completamente autonomo, in modo tale da ridurre al minimo il coinvolgimento umano e garantire una maggiore efficienza e velocità nel riconoscimento dei prodotti fraudolenti.

Indice

Elenco delle figure

Terminologia

Termine	Significato
Analisi del sentiment	Processo di analisi testuale per determinare se il tono di un testo scritto è positivo, negativo o neutro
API Key	Stringa alfanumerica univoca, generata da un algoritmo
CNP (Carta Non Presente)	Tipologia di frode in cui la carta fisica non è necessaria per la transazione
Cookie	Stringa univoca che identifica la sessione corrente su un browser
Data Point	Feedback positivo o negtivo in riscontro a una risposta formulata da un'IA per addestrarla
Endpoint	URL di un server o un servizio
e-WOM	Opinione diffusa tramite internet relativa a un prodotto, un servizio o un'azienda
Fine Tuning	Processo di addestramento di un modello di IA
GRU	Particolare tipo di rete neurale ricorrente
Job	Entità che rappresenta il prodotto durante tutta la fase di valutazione
LSTM (Long Short Term Memory)	Tipo particolare di architettura di rete neurale ricorrente
MCC (Matthews Correlation Coefficient)	Strumento statistico utilizzato per la valutazione dei modelli
NMT (Neural Machine Translation)	Metodologia di traduzione automatica che usa un modello statistico basato su apprendimento automatico
ORM (Object Relational Mapper)	Software che permette di interagire con database relazionali
Organizzazione	Azienda di e-commerce, denominata anche market place, registrata sull'applicativo
Query	Comando utilizzato per richiedere dati ad un database
RNN (Recurrent Neural Network)	Rete neurale in grado di elaborare dati sequenziali
Routing	Processo di selezione del percorso di una richiesta per far sì che venga servita correttamente
SDK (Software Development Kit)	Insieme di strumenti per lo sviluppo e la documentazione di software
SMOTE	Tecnica utilizzata nell'apprendimento automatico per trattare datasets sbilanciati
SVM	Algoritmo di apprendimento con supervisione
Webhook	Risultato della valutazione di un prodotto a seguito della sua validazione

Capitolo 1

INTRODUZIONE

Negli ultimi anni, l'e-commerce ha conosciuto un'espansione significativa, offrendo agli utenti un accesso senza precedenti a una vasta gamma di prodotti da tutto il mondo. Solo nel 2023, a livello mondiale l'e-commerce ha rappresentato oltre il 19% delle vendite al dettaglio e le previsioni indicano che per il 2027 costituirà quasi un quarto delle vendite al dettaglio globali[4]. Fondamentale per mantenere questo trend di crescita sarà preservare nel tempo la fiducia (trust) tra tutti i soggetti coinvolti: cliente, venditore e marketplace. Visto gli elevati controvalori finanziari in gioco, il rischio di frodi e comportamenti fraudolenti è sempre presente e rappresenta una minaccia costante per il settore. La fiducia è un elemento complesso da raggiungere e mantenere, e nell'ambito dell'e-commerce è strettamente legata da un lato, alla sicurezza delle transazioni finanziarie e dall'altro, alla qualità e affidabilità dei prodotti e dei servizi offerti.

La presente tesi tratterà in particolare, lo sviluppo di un'applicazione innovativa progettata per affrontare questo problema cruciale. L'obiettivo è quello di utilizzare l'intelligenza artificiale (IA) per identificare e valutare prodotti potenzialmente illegali o fraudolenti prima che questi vengano inseriti resi disponibili all'interno dei marketplace degli e-commerce. Questo approccio sfrutta il potenziale dell'IA, non solo per automatizzare il processo di verifica, ma anche per migliorare la sicurezza complessiva delle piattaforme di shopping online, riducendo drasticamente la possibilità di incorrere in truffe o vendita di prodotti o servizi illeciti.
Tutto ciò, viene fatto attraverso un portale web che permette, grazie ad un'interfaccia semplice e intuitiva, di visualizzare lo stato di avanzamento del processo di verifica per ogni singolo prodotto caricato e di modificare il risultato della verifica, restituendo un feedback al modello per migliorare la sua precisione e accuratezza.

Lo sviluppo dell'applicativo è iniziato in occasione del tirocinio curricolare svoltosi

presso l'azienda *Soluzioni Futura Srl* di Reggio Emilia, per poi proseguire anche dopo la fine dello stage. Il progetto è nato a seguito di una richiesta avanzata da un'azienda cliente della *Soluzioni Futura Srl* alla ricerca di un sistema in grado di rilevare automaticamente prodotti illegali o fraudolenti e che fosse più performante della soluzione da essi utilizzata sino a quel momento.

La necessità di questo progetto risiede nella crescente complessità degli acquisti online e nella diffusa presenza di prodotti contraffatti o illegali. Gli acquirenti spesso si trovano a rischiare l'acquisto di beni non conformi alle normative o potenzialmente dannosi. L'implementazione di un'applicativo che sfrutta le capacità predittive delle intelligenze artificiali offre una soluzione innovativa per mitigare tali rischi.

L'applicativo sviluppato rappresenta, quindi, uno strumento fondamentale per illustrare il potenziale delle intelligenze artificiali nel migliorare la sicurezza dell'e-commerce. Al di là della sua implementazione pratica, questa tesi si propone di sottolineare il ruolo cruciale delle IA nel settore della sicurezza informatica applicata al commercio digitale, evidenziando il potenziale trasformativo delle intelligenze artificiali nel garantire ambienti di shopping online più sicuri e affidabili per tutti i consumatori.

Capitolo 2

Contesto e stato dell'Arte

2.1 Intelligenza Artificiale

In questo capitolo si tratteranno alcune conoscenze preliminari legate all'intelligenza artificiale per comprendere al meglio tutte le caratteristiche dell'applicativo sviluppato per il riconoscimento di prodotti fraudolenti.

L'intelligenza artificiale (IA) è una delle discipline più dinamiche e trasformative nel panorama tecnologico odierno. Si riferisce alla capacità delle macchine di eseguire compiti che, se svolti dagli esseri umani, richiederebbero intelligenza. Questi compiti includono, ma non sono limitati a, l'apprendimento, il ragionamento, la risoluzione di problemi, la comprensione del linguaggio naturale e la percezione visiva.

Per i fini di questa tesi, è interessante approfondire i campi del **Natural Language Processing (NLP)**, dell'**Analisi testuale** e dell'**Anomaly Detection**.

2.1.1 Natural Language Processing

Il Natural Language Processing (NLP) è una sottodisciplina dell'IA che si occupa dell'interazione tra computer e linguaggio umano. Questa tecnologia permette ai sistemi di comprendere, interpretare e generare il linguaggio naturale in modo utile e significativo. Grazie a modelli avanzati come BERT [6] e GPT [25], il NLP è in grado di eseguire compiti complessi come la traduzione automatica, l'analisi del sentiment e la generazione di testo.

Ivano Lauriola [12] suddivide il NLP in due categorie principali: **ricerca fondamentale** (o di base), che include modelli linguistici, analisi morfologica, sintassi e analisi semantica; e **ricerca applicativa**, che si occupa di estrazione automatica di

informazioni, traduzione, sintesi, risposta automatica a domande, classificazione e clustering di documenti.

Tecniche e modelli utilizzati

1. **Deep Learning nel NLP**
    - **Modelli di Rete Neurale Profonda:** I recenti progressi nel deep learning hanno rivoluzionato il campo del NLP. Ad esempio, la traduzione automatica ha visto il passaggio dai modelli statistici, basati su frasi, ai modelli di traduzione automatica neurale (NMT[22]), che utilizzano reti neurali profonde per ottenere migliori performance.
    - **Reti Neurali Ricorrenti (RNN[14]) e Architetture Basate su Transformer:** Le architetture ricorrenti, come le Long Short-Term Memory (LSTM[8]) e le Gated Recurrent Units (GRU[21]), sono state ampiamente utilizzate per compiti di sequenza. Più recentemente, i Transformer hanno rivoluzionato il campo grazie alla loro capacità di gestire parallelamente l'attenzione tra tutte le parole di una sequenza, migliorando l'efficienza computazionale e le performance.
2. **Rappresentazione di parole e frasi**
    - **Word Embeddings:** La rappresentazione delle parole come vettori densi in uno spazio a bassa dimensione, introdotta da modelli come Word2Vec (vedi Fig. 2.1) e GloVe[16], ha permesso di catturare relazioni semantiche tra le parole. Questo è stato un punto di svolta nel NLP, permettendo alle parole con significati simili di avere rappresentazioni vicine nello spazio vettoriale.
    - **Sentence Embeddings:** Metodi come Doc2Vec e Skip-Thought[11] hanno esteso l'idea degli embeddings alle frasi, cercando di catturare il significato delle frasi intere invece che delle singole parole.
3. **Modelli Sequence-to-Sequence**
    - **Modelli Encoder-Decoder:** Utilizzati principalmente nella traduzione automatica e nella sintesi di testo, questi modelli comprendono un encoder che processa la sequenza di input e un decoder che genera la sequenza di output (Vedi Fig. 2.2). Il contesto dell'input viene compresso in un vettore di contesto, che viene poi utilizzato dal decoder per generare la traduzione.
    - **Meccanismi di Attenzione:** Introdotti per migliorare i modelli sequence-to-sequence, i meccanismi di attenzione permettono al modello di foca-

lizzarsi su parti specifiche della sequenza di input durante la generazione della sequenza di output, migliorando significativamente le performance, soprattutto per frasi lunghe.

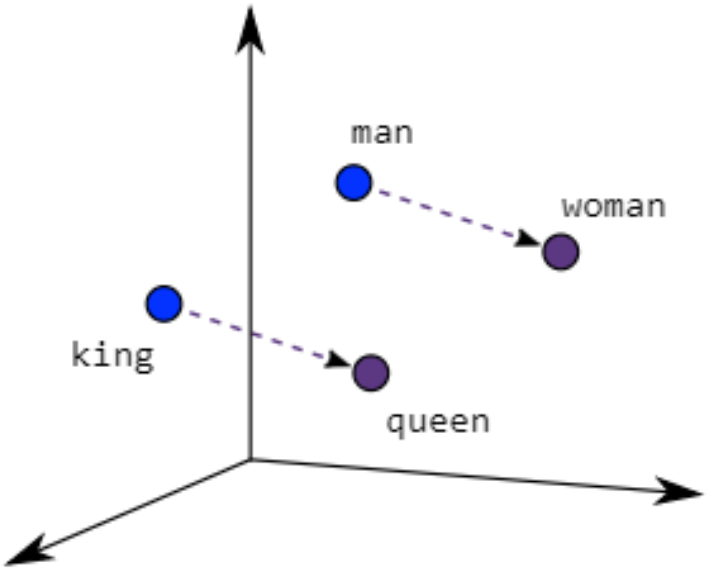

Figura 2.1: Esempio di parole relazionate semanticamente in un modello Word2Vec

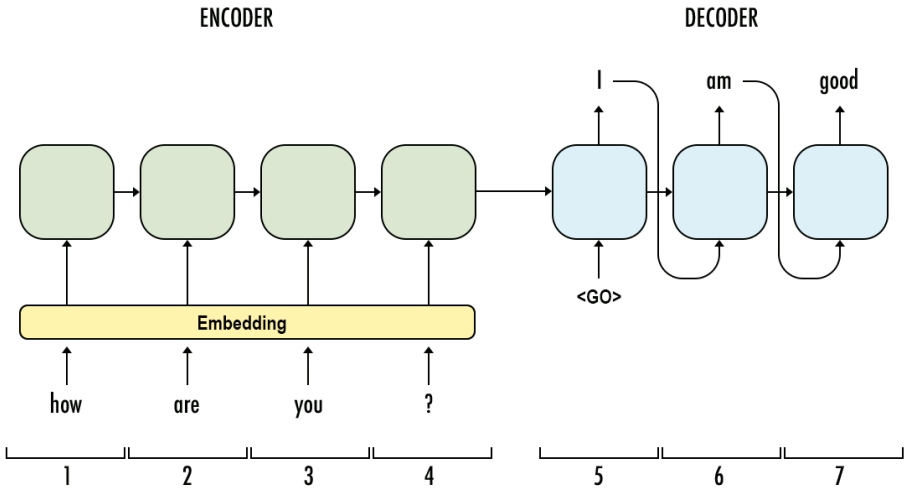

Figura 2.2: Schema di un modello Sequence-to-Sequence che utilizza un Encoder-decoder per interpretare una domanda e dare una risposta

2.1.2 Analisi Testuale

L'analisi testuale è il processo di estrazione di informazioni significative dai dati testuali. È un campo ampio che include tecniche come l'estrazione di informazioni e l'analisi del sentiment, che sono anche componenti chiave del NLP. Mentre il Natural Language Processing (NLP) si concentra sull'interazione tra computer e linguaggio umano, l'analisi testuale si occupa specificamente dell'interpretazione e della comprensione del contenuto testuale.

Per comprendere come l'intelligenza artificiale riesce a estrapolare informazioni dal testo e interpretarle, è utile esplorare i passaggi chiave di questo processo:

- Pre-Processing del testo
- Applicazione di tecniche di Text Mining
- Analisi delle informazioni estratte

Pre-Processing del testo

Il pre-processing del testo è una fase fondamentale per preparare i dati testuali all'analisi. E' il primo passo per estrarre informazioni significative dai testi e comprendere il loro contenuto. Questa fase include diverse operazioni, tra cui: estrazione, eliminazione delle stop-words, stemming e TF/IDF (Fig. 2.3).

Estrazione L'estrazione è il processo di suddivisione del testo in singole parole per analizzarle in modo più dettagliato. Questo passaggio è essenziale per creare una rappresentazione strutturata del testo e identificare le parole chiave.

Eliminazione delle stop-words Le stop-words sono parole comuni che non aggiungono significato al testo, come articoli, preposizioni, pronomi e congiunzioni. L'eliminazione delle stop-words è importante per concentrarsi sulle parole chiave e ridurre il rumore nei dati.
Esistono differenti metodologie per eliminare queste parole, come l'utilizzo di liste predefinite di stop-words o l'analisi delle frequenze delle parole per identificare quelle più comuni o quelle che compaiono solo una volta nel testo.

Stemming Lo stemming è il processo di riduzione delle parole alla loro radice, rimuovendo prefissi e suffissi per ottenere la forma base della parola. Questo passaggio è utile per ridurre le variazioni delle parole e semplificare l'analisi del testo.

TF/IDF Term Frequency-Inverse Document Frequency è una tecnica utilizzata per valutare l'importanza di una parola in un documento rispetto all'intero corpus.

Questo metodo assegna un peso a ciascuna parola in base alla sua frequenza nel documento e alla sua rarità nel corpus. Le parole che compaiono frequentemente in un documento ma raramente nel corpus avranno un punteggio più alto, indicando la loro importanza nel contesto specifico.

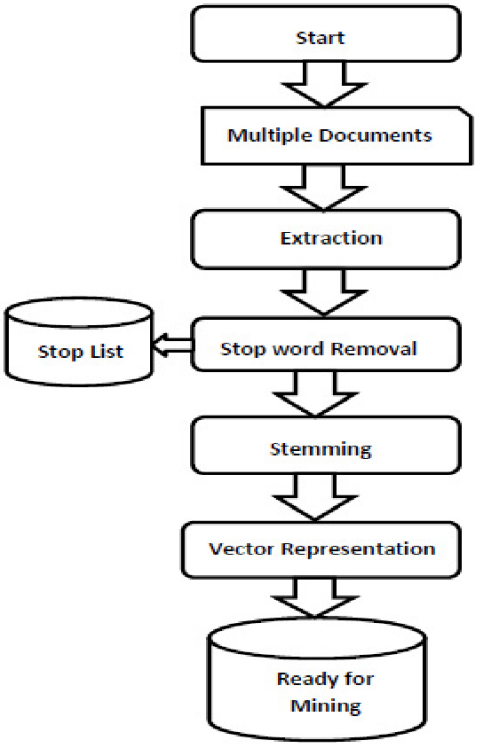

Figura 2.3: Schema rappresentante il percorso che segue un documento durante il pre-processing

Applicazione di tecniche di Text Mining

Dopo la fase di pre-processing vengono utilizzati strumenti e tecniche di text mining per analizzare i dati testuali e estrarre informazioni significative [23] [17]. Queste tecniche includono: estrazione di informazioni, categorizzazione, clustering e summerizing.

Estrazione di informazioni L'estrazione di informazioni è il processo di identificazione e cattura di informazioni rilevanti dai testi. Questo può includere l'identificazione di entità, relazioni e concetti chiave all'interno del testo.

Categorizzazione La categorizzazione è il processo di assegnazione di un'etichetta o una categoria a un documento in base al suo contenuto. Questo può essere utile per organizzare e classificare grandi quantità di dati testuali.

Clustering Il clustering è il processo di raggruppamento di documenti simili in cluster o gruppi. Questo può aiutare a identificare pattern e relazioni nei dati testuali e a creare una struttura organizzata per l'analisi.

Summarizing Il summarizing è il processo di creazione di un riassunto o una sintesi di un documento o di un insieme di documenti. Questo può aiutare a estrarre le informazioni chiave e a ridurre la complessità dei dati testuali.

Analisi delle informazioni estratte

Una volta che le informazioni sono state estratte dai testi, è possibile analizzarle per ottenere insight e informazioni significative. Questo step include l'interpretazione dei risultati, la visualizzazione dei dati e la generazione di report.

2.1.3 Anomaly Detection

L'anomaly detection, o rilevazione delle anomalie, è una tecnica utilizzata per identificare pattern anomali o comportamenti insoliti nei dati. Questo è particolarmente utile per rilevare frodi, errori o altre irregolarità nei dati. L'anomaly detection può essere applicata a diversi settori, tra cui la sicurezza informatica, la finanza, la sanità e il marketing.

Le anomalie possono essere di tipi diversi[2], come *point anomalies*, *contextual anomalies* o *collective anomalies*. Le *point anomalies* si riferiscono a punti di dati isolati che sono anomali rispetto al resto del dataset, come mostrato in Fig. 2.4, le *contextual anomalies* si verificano quando un punto di dati è anomalo in un contesto specifico ma non in un altro, mentre le *collective anomalies* si riferiscono a un insieme di punti di dati che insieme sono anomali, ma che presi singolarmente non lo sono.

Per identificare le anomalie nei dati, le tecniche che vengono utilizzate possono essere eseguite in tre modalità:

- **Supervisionata:** Questo approccio richiede un dataset di addestramento etichettato con anomalie e dati normali. Un modello di machine learning viene addestrato su questo dataset per identificare le anomalie nei nuovi dati.

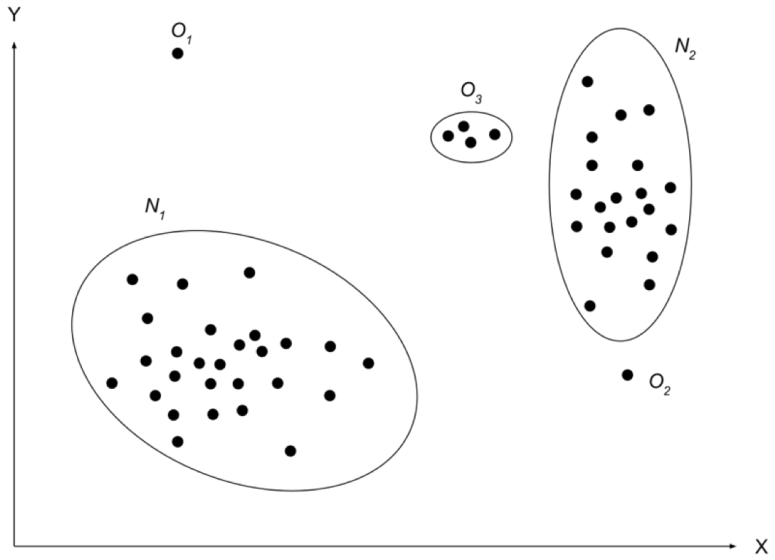

Figura 2.4: Rappresentazione in due dimensioni di casi di Point Anomalies

- **Non Supervisionata:** In questo caso, il modello viene addestrato sui dati senza etichette e cerca di identificare i pattern anomali nei dati. Questo approccio è utile quando non si dispone di dati etichettati.
- **Semi-Supervisionata:** Questo metodo combina elementi dei due precedenti, utilizzando un piccolo dataset etichettato e un dataset più grande senza etichette per identificare le anomalie.

Tecniche di Anomaly Detection

In questa sezione verranno presentate alcune delle architetture e tecniche più comuni[15] utilizzate per l'anomaly detection:

Metodi Statistici I metodi statistici sono spesso utilizzati per identificare anomalie nei dati. Questi consistono nell'analisi delle distribuzioni dei dati e nell'identificazione dei punti di dati che si discostano significativamente da queste distribuzioni. Alcuni esempi di metodi statistici includono il calcolo della deviazione standard, il calcolo della distanza euclidea e il calcolo della densità dei dati.

Metodi basati sul Machine Learning Questi metodi includono algoritmi di clustering, alberi decisionali, reti neurali e support vector machines. Questi modelli possono essere addestrati su dati etichettati o non etichettati per identificare le anomalie nei dati, tuttavia, a differenza degli approcci statistici che tendono a

concentrarsi sulla comprensione del processo che ha generato i dati, le tecniche di machine learning si concentrano sulla costruzione di un sistema che migliora le sue prestazioni basandosi sui risultati precedenti.

Metodi basati sul Data-Mining Con questi metodi, dopo aver pre-processato i dati e aver utilizzato tecniche di data mining, si cerca di identificare pattern anomali all'interno dei dati strutturati. Questo approccio è particolarmente utile per identificare anomalie in grandi dataset con molte variabili ed è funzionale anche quando non si hanno dataset etichettati.

Rule based Methods Questi metodi si basano su regole predefinite per identificare anomalie nei dati. Le regole possono essere definite manualmente o tramite l'analisi dei dati e possono essere utilizzate per identificare pattern anomali nei dati. Se si conoscono in anticipo i vincoli e le regole che definiscono cosa costituisce un'anomalia, questo approccio può essere particolarmente efficace.

2.2 Modelli e strumenti per il riconoscimento di frodi in ambito e-commerce

Diversi studi sono stati svolti per affrontare il problema delle frodi nell'ambito dell'e-commerce.
La letteratura esistente sull'argomento si concentra prevalentemente sul rilevamento e il riconoscimento di due macro-categorie di frodi: **transazioni finanziarie fraudolente** e **comportamenti disonesti degli utenti**, entrambe atte a danneggiare le piattaforme di e-commerce coinvolte, causando perdite finanziarie e danneggiando la reputazione delle aziende.

2.2.1 Trust Evaluation

Come anticipato nell'introduzione, l'importanza della valutazione della fiducia (trust) nei confronti degli e-commerce è fondamentale. Secondo uno studio condotto da Sofik Handoyo[10], i principali fattori che influenzano il comportamento d'acquisto online sono:

Fiducia: La fiducia è identificata come il fattore più influente nelle decisioni di acquisto online. Pertanto, le piattaforme di e-commerce dovrebbero concentrarsi strategicamente sulla costruzione e il mantenimento della fiducia dei consumatori. Il rafforzamento della fiducia può migliorare significativamente le vendite.

Rischio percepito: Il rischio percepito è importante nel processo decisionale dei consumatori online. Una gestione efficace del rischio percepito può aumentare la fiducia dei clienti e, di conseguenza, le vendite.

Sicurezza percepita: La sicurezza percepita gioca un ruolo significativo nelle decisioni di acquisto online, anche se non è il fattore più influente. È importante per le piattaforme di e-commerce focalizzarsi costantemente sulla sicurezza per alleviare le paure dei consumatori riguardo a possibili minacce. Assicurare e migliorare la sicurezza percepita deve rimanere una priorità per le piattaforme di e-commerce.

e-WOM: L'e-WOM (electronic Word of Mouth) ha un ruolo fondamentale come influenza sociale e fonte di informazioni per i consumatori.

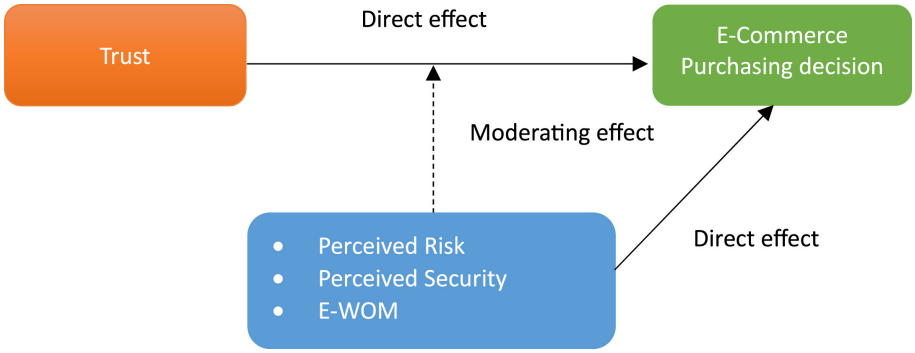

Figura 2.5: Fattori che influenzano direttamente e indirettamente l'intenzione delle persone ad acquistare online

Questi fattori sono strettamente correlati e interconnessi, e la loro gestione efficace può portare a un aumento della fiducia dei consumatori e delle vendite online, mentre una gestione inefficace può portare a una diminuzione della fiducia e dei profitti.

Riconoscimento di recensioni falsificate

Molti studi sono stati fatti a riguardo, e un primo di questi tratta l'importanza delle recensioni online. Secondo la ricerca condotta da Petr Hajek e Lubica Hikkerova[9] le recensioni online svolgono un ruolo fondamentale nella costruzione della fiducia dei consumatori nei confronti degli e-commerce.
Queste recensioni influenzano direttamente il comportamento d'acquisto, poiché i consumatori tendono a fidarsi maggiormente delle opinioni di altri acquirenti piuttosto che delle informazioni fornite direttamente dal venditore. Inoltre, le recensioni online sono considerate un'importante fonte di informazioni affidabili e imparziali, in grado di influenzare positivamente la percezione della qualità del prodotto e del servizio offerto dal venditore.

Attualmente, come mostrato nello studio, nei siti di e-commerce sono presenti numerose recensioni false, che influenzano negativamente la fiducia dei consumatori e la loro propensione all'acquisto.
Per affrontare questo problema, la ricerca propone l'implementazione di un modello avanzato di analisi del sentiment, noto come ABSA (Aspect-Based Sentiment Analysis). Questo modello, rispetto ai migliori esistenti, ha dimostrato una performance superiore nella rilevazione delle recensioni false. Il modello ABSA incorpora caratteristiche aggiuntive che permettono di identificare la manipolazione delle recensioni in modo più accurato, aumentando così l'informatività delle recensioni per le piattaforme di e-commerce.
In particolare, la considerazione delle informazioni legate al sentiment è cruciale per individuare recensioni sospette. La mancanza di tali informazioni rende difficile per i clienti e le piattaforme online rilevare le recensioni fasulle e prendere decisioni di acquisto informate. Pertanto, è consigliabile che le aziende che vendono prodotti e servizi online implementino meccanismi per monitorare gli aspetti legati al sentiment delle recensioni online e prendano l'iniziativa di segnalare eventuali attività di recensione dubbie.

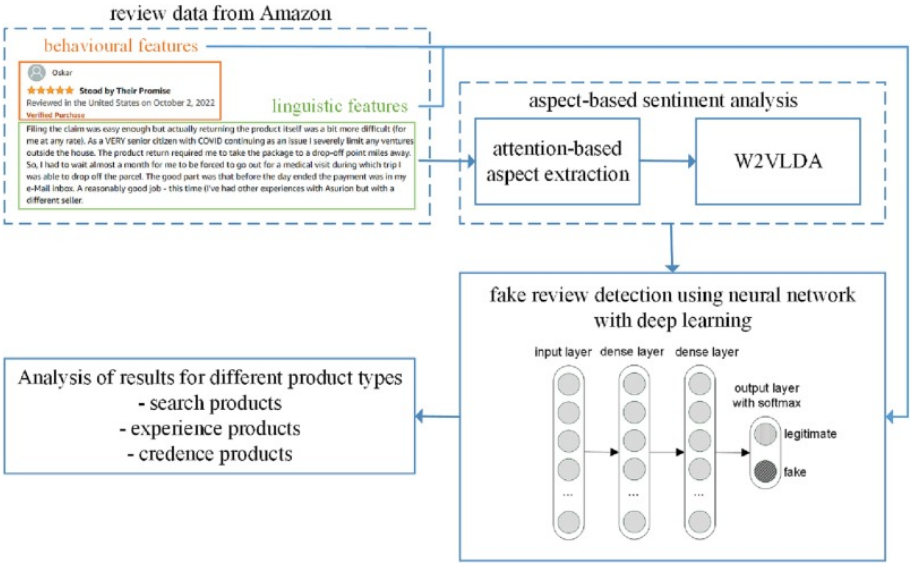

Figura 2.6: Schema dei componenti principali del modello ABSA per il riconoscimento di recensioni falsificate

Un altro studio interessante nell'ambito del riconoscimento delle recensioni false è quello condotto da Xin Li e Lirong Chen [13].
In questa ricerca, gli autori propongono di utilizzare una nuova caratteristica comportamentale dei recensori e una caratteristica periferica basata sulla fiducia indiretta, aggiungendo queste nuove variabili alle ricerche esistenti. Queste caratteristiche

sono poi combinate con le caratteristiche vettoriali pre-addestrate del modello linguistico BERT[6], creando un input per un modello di rete neurale.
Questo ha portato alla costruzione di un modello multi-modale per la rilevazione delle recensioni false, chiamato BMTBA (BERT + Multi feature + TextCNN + Bi-GRU + Attention). Questo modello è progettato per rilevare sia le recensioni falsi positivi che quelle negative.
Nonostante il modello sia stato testato principalmente su recensioni di ristoranti provenienti da diverse regioni su Yelp.com, gli autori suggeriscono che con un numero sufficiente di metadati, il modello potrebbe essere esteso anche ad altri settori.
I risultati delle prove dimostrano che il modello proposto supera significativamente il modello di riferimento in termini di accuratezza e altre metriche di valutazione, ottenendo eccellenti prestazioni nella rilevazione delle recensioni false.

2.2.2 Rilevamento di Transazioni Finanziarie Fraudolente

Le transazioni fraudolente rappresentano una minaccia significativa per le piattaforme di e-commerce, causando perdite finanziarie e danneggiando la reputazione delle aziende coinvolte. Come riportato da Praaven Kumar Sadineni[19] esistono diverse forme di transazioni fraudolente e le principali sono:

- **Skimming:** Una tipologia di frode in cui i dati della carta di credito o di debito vengono rubati tramite dispositivi elettronici illegali installati nei terminali di pagamento che acquisiscono i dati della carta per poterli clonare e utilizzare per effettuare transazioni non autorizzate.
- **Phishing:** Consiste nel tentativo di ottenere informazioni personali, come username, password e dettagli della carta di credito, attraverso l'utilizzo di e-mail o SMS, che portano l'utente a rivelare i propri dati sensibili.
- **Frode con carta non presente (CNP):** Questa tipologia di frode avviene quando i dati della carta di credito, dopo esser stati rubati, vengono utilizzati per effettuare transazioni online o per telefono che non richiedono la presenza fisica della carta stessa.
- **Furto d'identità:** Questa tipologia di frode avviene quando un individuo entra in possesso e utilizza le informazioni personali di un'altra persona, assumendone l'identità, per effettuare transazioni fraudolente, senza il consenso del legittimo proprietario.
- **Carta rubata durante la spedizione:** Consiste nel furto di carte di credito durante la spedizione, in cui i malintenzionati riescono ad ottenere le carte di credito prima che arrivino al legittimo proprietario.

- **Perdita della carta:** Questa tipologia di frode avviene quando la carta di credito viene persa dal legittimo proprietario e viene utilizzata da terzi per effettuare transazioni non autorizzate.
- **Siti falsi:** In questo caso i malintenzionati creano siti web falsi che imitano quelli delle aziende legittime per trarre in trappola gli utenti e ottenerene le informazioni personali e finanziarie.

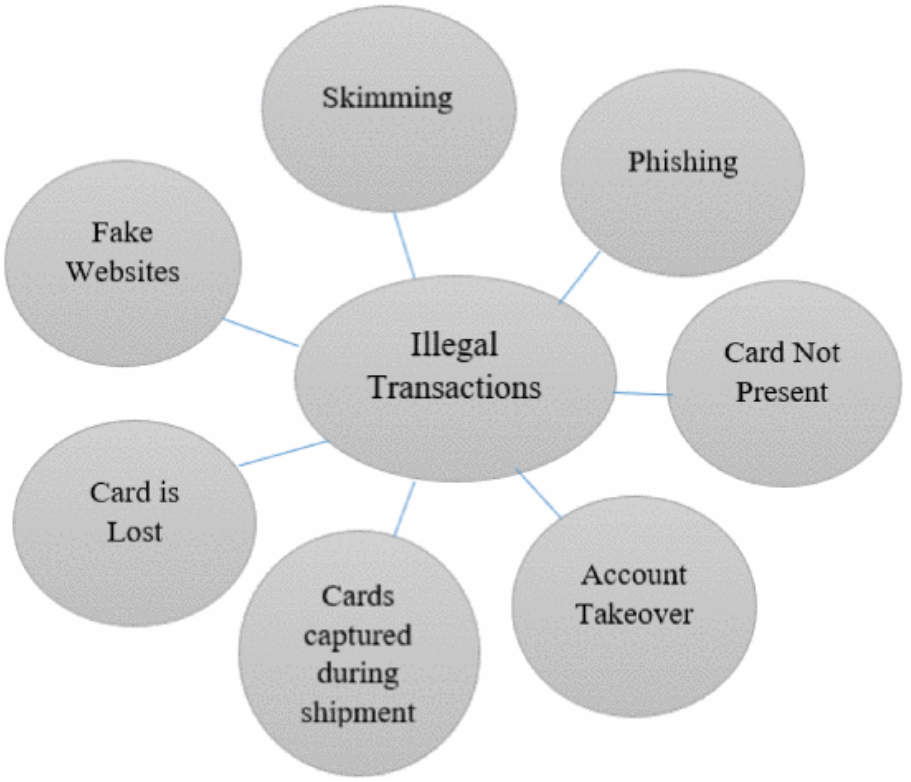

Figura 2.7: Schema delle principali tipologie di transazioni finanziare fraudolente

Per affrontare il problema delle transazioni fraudolente, diversi modelli di machine learning sono stati proposti per identificare e prevenire le frodi.

Vaishnavi Nath Dornadula[7] propone un metodo innovativo sviluppato per la rilevazione delle frodi, basato sull'analisi dei comportamenti dei clienti attraverso le loro transazioni. Questo metodo inizia raggruppando i clienti in base ai loro comportamenti di spesa, creando così un profilo dettagliato per ogni titolare di carta di credito. Successivamente vengono applicati diversi algoritmi di classificazione su tre diversi gruppi di transazioni e per ogni tipo vengono generati dei punteggi di valutazione. Questi cambiamenti dinamici nei parametri consentono al sistema di

adattarsi tempestivamente ai nuovi comportamenti di spesa dei titolari di carta di credito.

Inoltre, è stato implementato un meccanismo di feedback per affrontare il problema del *concept drift*, ovvero il cambiamento nel tempo dei comportamenti di spesa che può influenzare l'efficacia del rilevamento delle frodi. Durante lo studio, è stato osservato che il "Matthews Correlation Coefficient" (MCC[24]) era il parametro più efficace per gestire i dataset sbilanciati, dove le frodi sono molto meno frequenti rispetto alle transazioni legittime.

Tuttavia, l'MCC non è stata l'unica soluzione adottata. È stata utilizzata anche la tecnica SMOTE[3] (Synthetic Minority Over-sampling Technique) per bilanciare il dataset, riscontrando che i classificatori performavano meglio dopo questa operazione. Un altro approccio per gestire dataset sbilanciati è stato l'uso di classificatori a classe singola, come il *one-class SVM*[5].

Infine, lo studio ha evidenziato che gli algoritmi di regressione logistica, l'albero decisionale e la foresta casuale erano quelli che davano i migliori risultati nella rilevazione delle frodi.

Un altro studio riguardante il rilevamento di transazioni fraudolente è quello di Imane Sadgali[18]. In questo paper è stata confrontata l'efficacia di quattro tecniche di machine learning supervisionato per la rilevazione delle frodi con le carte di credito: l'albero decisionale (decision tree), il k-nearest neighbor, le foreste casuali (random forests) e le macchine a vettori di supporto (support vector machines, SVM).

Per la valutazione, è stato utilizzato un unico dataset generico di transazioni con carte di credito. Lo scopo di questo confronto era verificare se i risultati ottenuti fossero in linea con quelli della letteratura esistente, la quale spesso presenta risultati basati su dataset specifici. Dopo aver analizzato la performance di ciascuna tecnica, è emerso che le macchine a vettori di supporto (SVM) hanno dimostrato essere la tecnica più efficace tra quelle esaminate.

Capitolo 3

Una piattaforma per il riconoscimento di prodotti fraudolenti in siti di e-commerce con tecniche di IA

In questo capitolo verrà trattata la metodologia adottata per l'implementazione dell'applicativo web per il riconoscimento di prodotti fraudolenti in siti di e-commerce, senza entrare troppo nei dettagli tecnici, che verranno invece discussi nel capitolo successivo.

L'obiettivo principale della piattaforma è quello di interfacciarsi con intelligenze artificiali già esistenti e liberamente accessibili (GPT[25], Claude[1], Gemini[20], ecc...) per identificare e valutare prodotti potenzialmente illegali o fraudolenti prima che questi vengano inseriti all'interno dei cataloghi delle aziende di e-commerce.

3.1 Attori coinvolti

L'applicativo consente ad un insieme di utenti l'esecuzione di determinate azioni in base al proprio ruolo/profilo. Tutti gli utenti che fanno parte della medesima Organizzazione, hanno accesso all'intero catalogo dei prodotti caricati dall'Organizzazione stessa.
Esistono due tipi di ruoli per gli utenti:

1. **Operator:** Identifica gli utenti con ruolo caratterizzato dalle maggiori limitazioni. Permette l'accesso alla dashboard dei prodotti e alla pagina del proprio account personale.

2. **Admin:** Identifica gli utenti che hanno compiti di amministrazione e permette l'accesso a tutte le pagine dell'applicativo, incluse: la pagina di gestione dell'Organizzazione, la dashboard delle API keys e la dashboard dei webhooks.

Gli utenti di entrambi i ruoli possono validare manualmente ogni prodotto, ovvero possono confermare il risultato fornito dall'IA, se sono d'accordo, oppure rigettarlo se, invece, ritengono che l'IA lo abbia valutato erroneamente.

3.2 Percorso di validazione dei prodotti

Ogni prodotto caricato nell'applicativo, inizialmente viene elaborato e suddiviso in 3 Jobs (repliche identiche del prodotto originale), che vengono inviati separatamente allo stesso modello di IA per essere valutati. Una volta che tutti e 3 i Jobs sono stati valutati, si procede con la valutazione finale del prodotto, che viene considerato fraudolento se almeno 2 dei 3 Jobs sono stati valutati come tali. In caso contrario, il prodotto viene considerato lecito.
La tecnica appena descritta, permette di rappresentare l'incertezza dell'IA sulla valutazione di un determinato prodotto, in quanto se non tutti i Jobs hanno ricevuto lo stesso risultato, l'IA non è sicura al 100% ed è un dato fondamentale per l'utente che deve confermare o rigettare il risultato proposto. Al completamento di questa prima fase, viene eseguita un'ultima richiesta al modello di IA utilizzato fino a questo momento per l'elaborazione dei Jobs, per generare una nuova motivazione partendo dalle 3 motivazioni precedentemente fornite in risposta ad ognuno dei Job, che andrà a comporre il risultato finale che viene visualizzato sull'applicativo.
Dopo questo procedimento, il prodotto sarà validato automaticamente o manualmente (in base alla configurazione scelta dall'Organizzazione), mandando il feedback all'IA per l'addestramento del modello utilizzato e inviando il risultato della valutazione all'Organizzazione.

Di seguito, nella Fig. 3.1, è illustrato il processo appena descritto.
Si parte dal caricamento del prodotto da parte dell'Organizzazione e si termina con la validazione da parte di un utente, della valutazione data dall'IA relativa al prodotto, seguita dalla restituzione del risultato all'Organizzazione stessa e dall'invio del feedback all'intelligenza artificiale per permetterle di fare fine tuning.

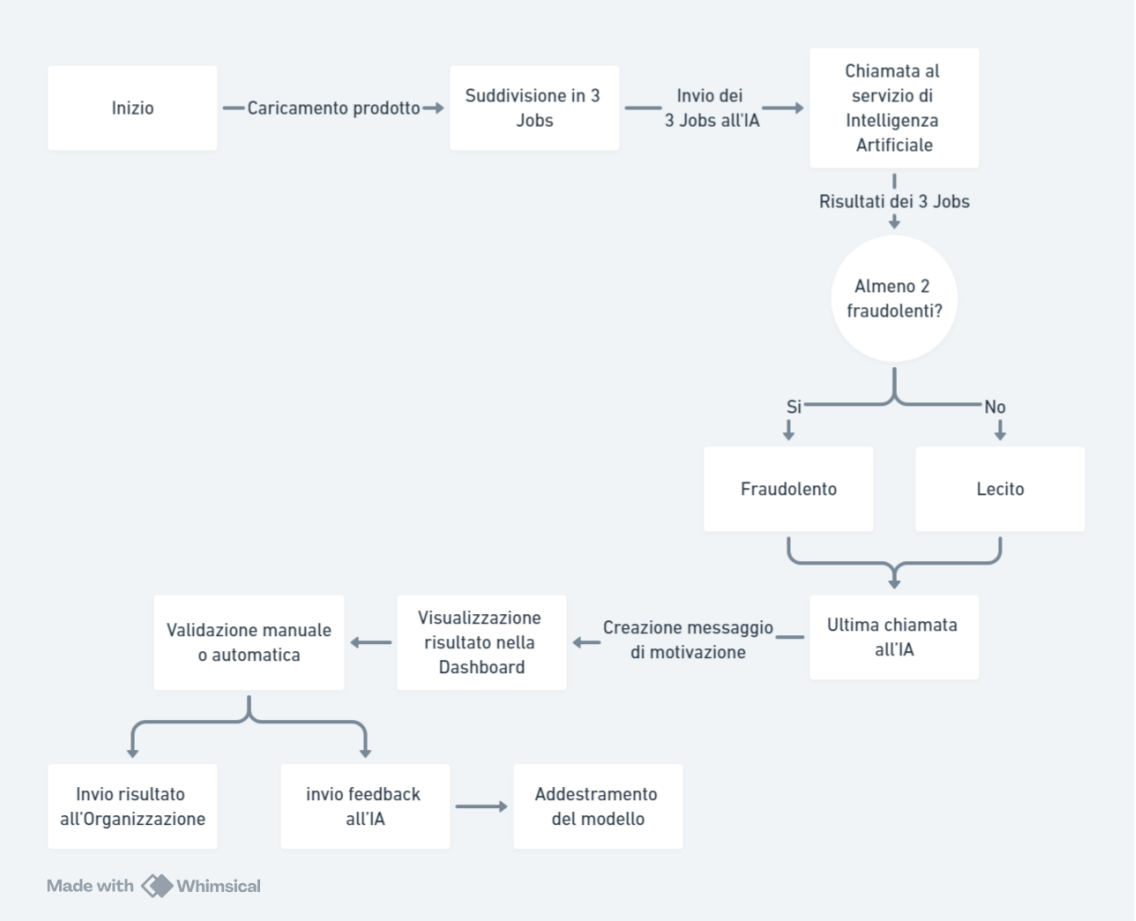

Figura 3.1: Rappresentazione del processo di valutazione e validazione di un prodotto, dal suo caricamento al suo completamento

3.3 Architettura dell'applicativo

In questa sezione verranno approfondite tutte le interfacce con cui l'utente può interagire e il collegamento tra il *frontend*, il *backend* e il *database*.
L'applicativo è in grado di gestire differenti Organizzazioni (tipicamente l'azienda di e-commerce che ne usufruisce) e relativi utenti, garantendo separatezza dei dati a livello di database tra differenti Organizzazioni.
È presente anche una console di amministrazione per la gestione delle Organizzazioni (creazione, modifica e cancellazione) e la gestione dell'utente ADMIN.

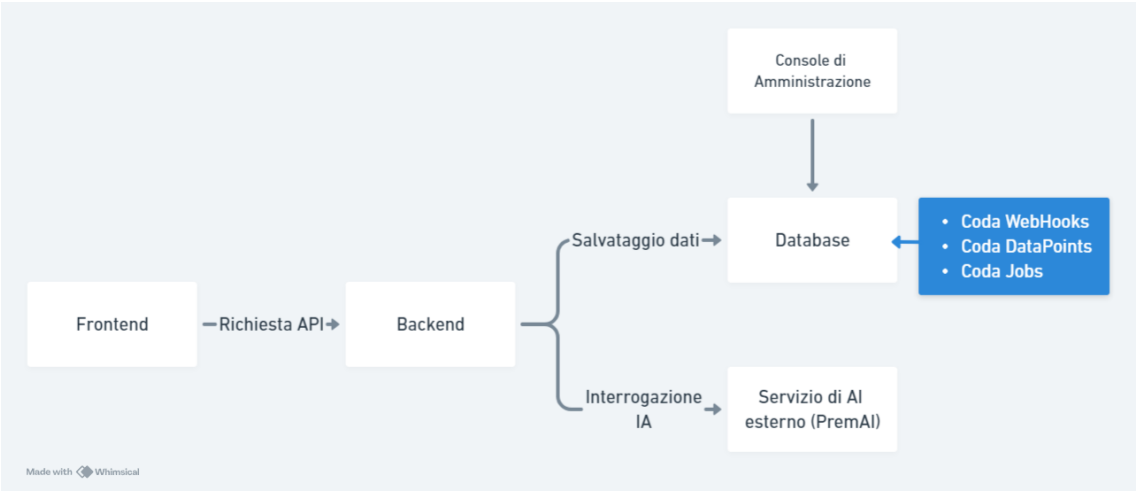

Figura 3.2: Schema architetturale dell'applicativo

3.3.1 Interfaccia Utente

La progettazione dell'interfaccia utente è stata condotta tenendo in forte considerazione la UX, al fine di renderla semplice e intuitiva.

Pagina Utente

Qui viene mostrata una panoramica generale delle informazioni personali dell'utente, come lo *username*, l'*email*, il *nome* e il *cognome*. Inoltre l'utente, da questa pagina, può modificare i propri dati e la password.

Dashboard dell'Organizzazione

In questa pagina è possibile visualizzare e gestire il nome e l'URL dell'endpoint esposto dall'Organizzazione per inviare i webhooks. Qui si possono anche gestire le preferenze dell'Organizzazione, come la percentuale di affidabilità minima per validare automaticamente un prodotto lecito o fraudolento e i criteri di valutazione che verranno aggiunti al prompt di default inviato all'IA ad ogni elaborazione. Ogni Organizzazione dovrà selezionare almeno 3 dei 6 criteri proposti, visibili in Fig. 3.3.

- **Prodotti ottenuti illegalmente:** Tutti i prodotti che è presumibile che non siano stati ottenuti legalmente, come le credenziali di un account bancario.
- **Prodotti che aiutano a eseguire attività illegali:** Prodotti che non sono illegali, ma che hanno lo scopo di aiutare ad eseguire delle azioni illegali. Un esempio è uno strumento per hacking.

- **Prodotti che non violano l'EULA:** Tutti i prodotti che devono essere valutati come illeciti anche se non violano espressamente l'EULA, come per esempio, i cheat per i videogiochi.
- **Prodotti che richiedono la spedizione:** Nel caso l'Organizzazione voglia solo vendere prodotti digitali, questo criterio permette di specificare che non devono essere validi dei prodotti concreti che richiederebbero una spedizione.
- **Prodotti non chiari:** Questo criterio serve per valutare come fraudolento un prodotto che non contiene abbastanza informazioni per determinare se è lecito o meno.
- **Prodotti venduti a un prezzo estremamente diverso rispetto al loro valore:** Tutti i prodotti che sono venduti con una differenza di prezzo molto elevata rispetto al valore di mercato e che, per questo motivo, presumibilmente sono fraudolenti. Un esempio può essere una licenza di Windows venduta a 2$.

I criteri selezionati sono, quindi, dei vincoli aggiuntivi che permettono all'IA di valutare come fraudolenti tutti i prodotti che rispettano quegli stessi criteri.

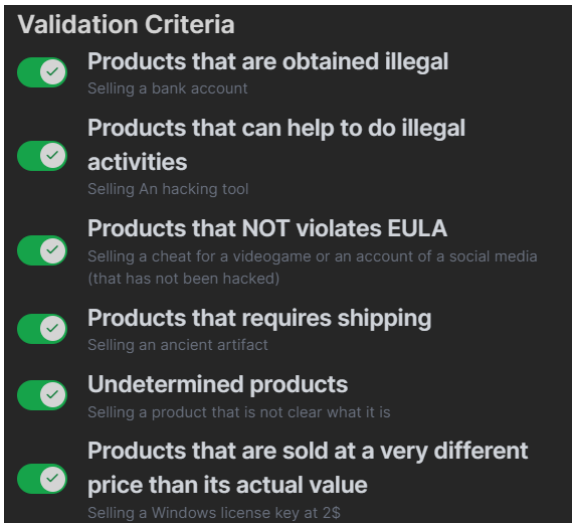

Figura 3.3: Lista dei sei criteri per personalizzare in base alle policy aziendali la valutazione dei prodotti

È inoltre presente una tabella che mostra tutti gli utenti appartenenti alla medesima Organizzazione, con le relative informazioni, quali lo *username*, l'*email*, il *nome*, il *cognome*, il *ruolo* e le *azioni permesse*. Da questa tabella è possibile eliminare un utente esistente o modificarne il ruolo.

Ogni utente che viene aggiunto all'Organizzazione deve essere invitato, e da questa pagina è possibile accedere alla dashboard degli inviti, dove è possibile visualizzarne lo stato.

Dashboard delle ApiKeys

In questa sezione è possibile visualizzare e gestire le **chiavi API**, necessarie per il caricamento dei prodotti. Ogni Organizzazione può disporre di più chiavi e da questa pagina può eliminarle o crearne di nuove.
Per ognuna vengono mostrati il nome e i primi e gli ultimi 5 caratteri della chiave, non permettendo di visualizzarla completamente per motivi di sicurezza.

Coda dei WebHooks

L'applicativo permette di visualizzare tutti i **webhooks** creati dall'Organizzazione, con le relative informazioni, come l'ID del prodotto a cui fa riferimento e lo stato del webhook. Nel caso in cui si dovessero verificare anomalie durante l'invio, verranno visualizzate anche la data e l'ora del prossimo tentativo automatico, con la possibilità di forzarne manualmente l'immediata esecuzione. Nel caso si tenti di modificare un webhook già inviato, verrà creato un nuovo webhook senza eliminare quello già esistente.

Dashboard Prodotti

Questa pagina è l'elemento centrale dell'applicativo. Come visibile in Fig. 3.4, in essa vengono mostrati tutti i prodotti caricati dall'Organizzazione, con le relative informazioni. Per ogni prodotto è indicato:

- **ID:** L'ID univoco del prodotto così come è identificato nella piattaforma di e-commerce che lo ha caricato.
- **Name:** Il titolo del prodotto.
- **Description:** La descrizione del prodotto.
- **Submission:** La data e l'ora in cui il prodotto è stato caricato.
- **AI Response:** Lo stato di avanzamento della valutazione del prodotto, che può assumere i seguenti valori:
    1. **Pending:** In attesa di essere processato.
    2. **Validating:** In corso di validazione.
    3. **Failed:** Il prodotto non è riuscito ad essere validato per problemi tecnici dovuti alla connessione o all'indisponibilità dell'IA.

4. **Valutato:** Il prodotto è stato valutato correttamente e in questo caso viene mostrato il risultato dell'analisi effettuata dall'IA, accompagnata da una percentuale di affidabilità.

- **Actions:** Qui si può confermare il risultato dell'IA oppure rigettarlo. Ogni volta che si esegue una di queste due azioni, o si crea un nuovo webhook o se ne modifica uno esistente, se non è stato già inviato. In ogni caso si restituisce un feedback al modello di IA per permettergli di fare fine tuning.
  Nel caso in cui il prodotto sia stato validato automaticamente, viene visualizzata un'icona apposita per notificare che non è stato validato manualmente, ma rimane comunque la possibilità di rigettare il risultato proposto dall'IA.

I prodotti sono suddivisi su più pagine (impaginazione) per facilitarne la consultazione e la gestione, e ogni volta che avviene una modifica ai prodotti stessi, l'utente viene notificato tramite un messaggio pop-up o, nel caso il prodotto sia all'interno della pagina che sta visualizzando, viene aggiornata la tabella in tempo reale.
La pagina inoltre permette di filtrare i prodotti in base al loro stato o di cercare un prodotto specifico per ID.

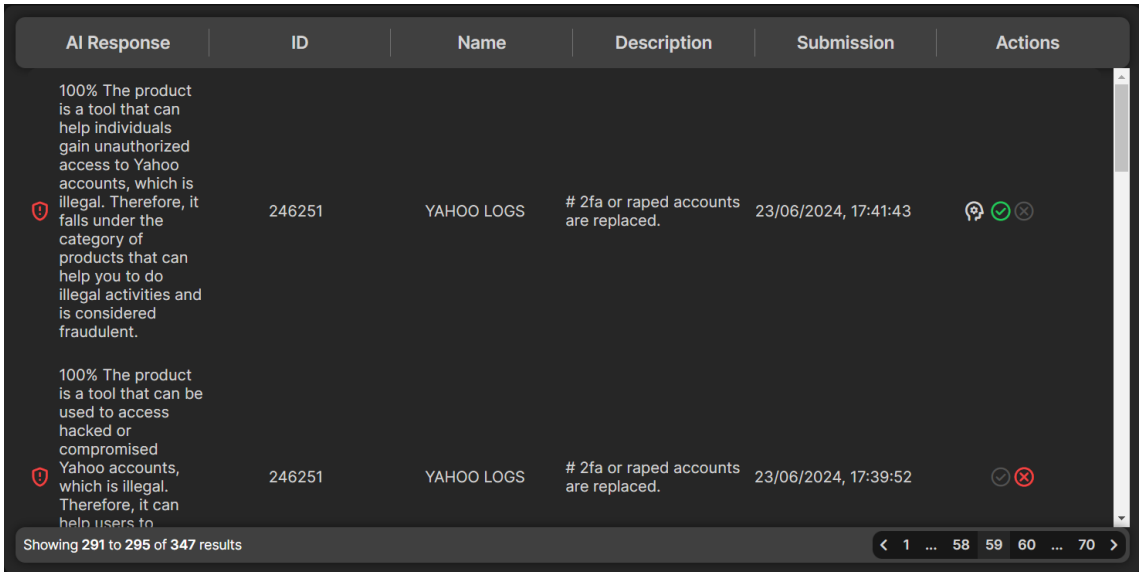

Figura 3.4: Tabella dei prodotti dove il primo è stato validato automaticamente, mentre il secondo manualmente

3.3.2 Backend

Il backend dell'applicativo è progettato per garantire l'integrità dei dati e la sicurezza. Gestisce la sessione di ogni utente loggato e si occupa di reindirizzare tutte

le richieste ai vari endpoint che servono per offrire i servizi all'interfaccia web.
Il backend si appoggia su un database per memorizzare i dati degli utenti, delle Organizzazioni, dei prodotti e delle valutazioni effettuate dall'IA.

Funzionalità Chiave

Routing Ogni tipologia di servizio offerto dall'applicativo è raggiungibile tramite un URL specifico, che viene chiamato dall'interfaccia web. Tutti gli endpoint sono protetti da un sistema di autenticazione che permette di verificare l'identità dell'utente che sta effettuando la richiesta e di garantire che solo gli utenti autorizzati possano accedere ai servizi offerti dall'applicativo, tranne per i servizi di registrazione, di login e per le pagine di errore.

Gestione del Database Il database dell'applicativo è progettato per memorizzare i dati in modo sicuro e organizzato. Ogni tabella del database è progettata per salvare un tipo specifico di dato, come gli utenti, le Organizzazioni, i prodotti e le valutazioni effettuate dall'IA.
Lo scopo principale del database è garantire la persistenza dei dati per permettere dei retry in caso di fallimento di ogni tipo di richiesta interna o esterna, dall'invio dei webhook alla richiesta di validazione dei prodotti da parte dell'IA. Questo è fatto con un sistema di code nelle quali vengono salvati tutti i dati che devono essere inviati e che, in caso di fallimento, non possono essere persi.

Scripts di controllo

Una parte fondamentale del backend è composta da alcuni script che si occupano di controllare ciclicamente alcuni elementi all'interno del database.

1. **WebHookSender:** Questo script si occupa di controllare la coda dei webhooks sul database e di inviare i webhooks che sono pronti.
2. **DataPointSender:** Questo script si occupa di inviare i data points all'IA per permetterle di fare fine tuning.
3. **ProductValidator:** Questo script si occupa di controllare i Jobs in attesa di essere validati, dando precedenza a quelli in stato di errore e di inviarli all'IA per essere valutati.

Utilizzando questi script, che vengono eseguiti in parallelo, si garantisce che l'applicativo sia sempre aggiornato e che i dati siano sempre consistenti.

3.4 Requisiti Funzionali

3.4.1 Login e Sign-Up

Queste due pagine sono fondamentali per gestire la sessione dell'utente durante l'utilizzo dell'applicativo. La pagina di login permette di autenticarsi per accedere al portale. Nel caso si cerchi di accedere in modo diretto ad altre pagine senza aver effettuato l'accesso, si viene reindirizzati alla pagina di autenticazione.
Alla pagina di Sign-up si può accedere solo previo invito da parte di un utente Admin dell'Organizzazione, infatti per ogni invito viene creato un link univoco, valido per una settimana, senza il quale non è possibile registrarsi.

3.4.2 Caricamento dei Prodotti

L'applicazione è in grado di ricevere i dati relativi ai singoli prodotti da valutare, esponendo un endpoint che permette all'Organizzazione che usufruisce della piattaforma, di farli valutare prima ancora che questi vengano resi disponibili sul proprio marketplace.
I dati relativi al prodotto includono informazioni come il titolo e la descrizione, e una volta importati, vengono man mano processati dall'IA per determinare se il prodotto è potenzialmente fraudolento o illegale.

3.4.3 Valutazione del Prodotto

Per la valutazione tramite IA, l'applicativo si appoggia ad un servizio esterno che permette, tramite API, di interfacciarsi direttamente a diverse intelligenze artificiali. Ad ogni prodotto viene affiancato un prompt di default che descrive come deve comportarsi l'IA per valutare il prodotto, definisce i vincoli e le regole da seguire e specifica quali informazioni sono rilevanti per la valutazione. In aggiunta, l'Organizzazione può personalizzare il prompt con delle specifiche preesistenti, in modo da adattarlo alle proprie esigenze e alle proprie policy.

3.4.4 Portale di Gestione dei Prodotti

L'applicativo mette a disposizione un'interfaccia web per la gestione dei prodotti, dove gli utenti possono visualizzare lo stato di avanzamento della valutazione di ciascun prodotto, confermare o rigettare singolarmente i risultati dell'IA e visualizzare i dettagli delle valutazioni effettuate.

3.4.5 Fine Tuning dell'IA

Ogni volta che un utente interagisce con l'applicativo per confermare o rigettare un risultato proposto dall'IA, viene inviato un **Data Point** al servizio esterno che permette di raffinare e migliorare l'efficacia dell'IA. Questo processo di fine tuning permette al modello di apprendere dai feedback ricevuti e di migliorare le proprie prestazioni nel tempo (autoapprendimento).

3.4.6 Invio dei Risultati

Quando un prodotto viene validato manualmente o automaticamente, si aggiunge un webhook alla coda di invio, che notifica l'Organizzazione del risultato della valutazione. Il webHook viene inviato trascorsi 5 minuti dalla sua creazione, lasciando quindi il tempo per correggere eventuali errori umani o per rigettare una validazione automatica prima che venga restituito all'Organizzazione.

3.5 Integrazione con le Intelligenze Artificiali

I modelli di IA ricoprono un ruolo fondamentale per il corretto funzionamento dell'applicativo, in quanto si occupano di svolgere l'attività core: la valutazione dei prodotti.
Per interfacciarsi con i vari modelli di intelligenza artificiale, l'applicativo si appoggia ad un servizio esterno: **PremAI**. Questo servizio permette, tramite delle semplici chiamate API, di connettersi a uno dei tanti modelli messi a disposizione, dando la possibilità di integrare un sistema di interrogazione di intelligenze artificiali senza doverlo implementare.
I modelli più performanti e più utilizzati di cui dispone sono Claude, Gemini e GPT. Ognuno di questi in più varianti come, nel caso di GPT, *GPT-3.5*, *GPT-4* e *GPT-4o*. Proprio quest'ultima, è stata quella prescelta, in quanto è risultata essere la più performante e la più adatta alle esigenze. Esistono molti altri modelli disponibili su PremAI, come *Mistral*, *Llama* o *Command R*, ma molti di loro non si sono dimostrati sifficientemente efficaci per il compito richiesto.

3.5.1 Prompt di default

Ogni volta che viene interrogata l'IA per la valutazione di un prodotto, assieme a questo viene inviato un prompt specifico che è stato creato dopo un'attenta analisi delle regole e dei vincoli necessari per permettere all'IA di comportarsi al meglio.
Il prompt è diviso in più punti, ognuno dei quali rappresenta una categoria di vincolo che l'IA deve rispettare:

1. **Role:** Definisce il ruolo nel quale deve immedesimarsi per valutare il prodotto.
2. **Task:** Definisce il compito che deve svolgere.
3. **Criteria:** Insieme di regole e vincoli che rappresentano regole comuni. (Es. *"Products that can help you to do illegal activities are fraudolent even if they are not illegal themselves"*)
4. **Additional Rules:** Regole aggiuntive e fondamentali per la valutazione del prodotto. (Es. *"YOU MUST HAVE A CLEAR CASE OF FRAUD TO FLAG A PRODUCT AS FRAUDOLENT"*)
5. **Examples:** Alcuni esempi per aiutare l'IA a capire meglio il contesto in cui si trova.
6. **Input:** Il formato in cui vengono inviati dati al modello.
7. **Output:** Il formato in cui vengono restituiti i dati dal modello.
8. **Process:** Il processo, passo per passo, che l'IA deve seguire per valutare il prodotto. Questo include la generazione della motivazione per la quale il prodotto è stato valutato in un certo modo prima di prendere una decisione, che permette al modello di raggiungere una decisione più accurata.

Capitolo 4

Implementazione dell'applicativo

In questo capitolo verranno trattati i dati più tecnici dell'applicativo, come l'architettura, le tecnologie utilizzate e le scelte progettuali fatte per garantire il funzionamento dell'applicativo.

4.1 Architettura

4.1.1 Frontend

L'interfaccia dell'applicativo è stata costruita in HTML, utilizzando la libreria TailwindCSS[1] per la componente grafica e la libreria HTMX[2] per la gestione efficace delle richieste HTTP.

TailwindCSS

TailwindCSS è un framework CSS che permette di creare interfacce web personalizzate e moderne. Utilizza una metodologia utility-first, che permette di creare stili personalizzati e flessibili senza dover scrivere CSS personalizzato.

HTMX

HTMX è una libreria JavaScript che permette di aggiungere funzionalità dinamiche alle pagine web senza dover scrivere codice JavaScript personalizzato. Questa libreria permette di definire determinate azioni e opzioni, semplicemente definendole come attributi all'interno di un tag HTML.

[1] https://tailwindcss.com/docs/installation
[2] https://htmx.org/docs/

La motivazione che ha portato alla scelta di HTMX è stata quella di voler lasciare il più possibile la logica lato server per rendere il client semplice e leggero. Ogni risposta che arriva dal server non è altro che un pezzo di HTML opportunamente creato e formattato per essere inserito direttamente e senza ulteriori modifiche all'interno della pagina web.

HTMX all'interno del progetto è stato utilizzato per due scopi principali:

- **Gestione delle richieste AJAX:** HTMX permette di effettuare richieste HTTP definendo il metodo della chiamata, l'url della destinazione, il contenuto da inviare e l'elemento html destinatario nel quale verrà scritta la risposta.
- **Server Sent Events:** HTMX implementa funzionalità predefinite che permettono di usufruire degli eventi inviati dal server in tempo reale. Questo approccio è stato utilizzato per aggiornare tutte le tabelle presenti nell'applicativo.

4.1.2 Backend

La parte più complessa dell'applicativo, vista la scelta progettuale di adottare la libreria HTMX, è il backend. Quest'ultimo è stato sviluppato in NodeJS, utilizzando il framework Fastify [3] per il routing delle richieste e un ORM per la comunicazione efficiente con il database.

NodeJs

Node.js è un ambiente di runtime JavaScript open-source e multipiattaforma che permette di eseguire codice JavaScript lato server. È basato sul motore JavaScript V8 di Google Chrome e permette di creare applicazioni web scalabili e performanti.

Fastify

La scelta di Fastify è stata dettata dalla sua velocità e limitato consumo di risorse. Fastify è un framework web per Node.js che permette di creare applicazioni web veloci e scalabili. È stato progettato per essere leggero e performante, offrendo un'alternativa più veloce rispetto ad altri framework più diffusi come Express.
Sono state implementate diverse route, ognuna per gestire un determinato servizio offerto dall'applicativo, come la gestione degli utenti, delle organizzazioni, dei prodotti e dei webhooks.

[3] https://fastify.dev/docs/latest/

SSE (Server Sent Events)

Lato client, si utilizza HTMX per gestire gli eventi, mentre lato server viene utilizzata una libreria chiamata SSEManager, creata in-house, per inviarli. L'invio degli eventi è gestito tramite un sistema di stanze, alle quali, ogni client si può iscrivere per riceverli (es. quando si visita la dashboard dei prodotti viene aperta una connessione http, ad uno specifico endpoint, che non viene chiusa e che fa iscrivere il client a una stanza). Una volta che un evento viene generato, viene inviato in broadcast all'interno di una stanza specifica. Questo sistema è utilizzato per aggiornare in tempo reale le tabelle, inviando il contenuto della riga che viene aggiornata e che grazie ad HTMX viene sostituita senza bisogno di ricaricare la pagina.

Database

Il database utilizzato è relazionale e si basa su PostgreSQL[4].
Ogni tabella è stata progettata per memorizzare un tipo specifico di dato, come gli utenti, le organizzazioni, i prodotti e le valutazioni effettuate dall'IA. Le tabelle sono collegate tra loro tramite chiavi esterne per garantire l'integrità dei dati e permettere di effettuare ricerche e analisi complesse.
Il collegamento con il database è stato sviluppando utilizzando un ORM chiamato Drizzle[5], che rende molto più semplice l'interrogazione al database e la gestione delle relazioni tra le varie entità.

L'approccio all'accesso del database è stato gestito in maniera diversa dal metodo classico. Infatti per ridurre al minimo le query, così da ridurre anche i possibili errori di connessione, quando si richiede una risorsa, si riceve, non solo la risorsa stessa, ma tutte quelle ad essa collegate (es. eseguendo una query che richiede un prodotto specifico, non solo si ottiene il prodotto, ma anche l'Organizzazione al quale appartiene, i Job creati a partire da quel prodotto e i risultati dei singoli Job). Tutto questo è possibile perché PostgreSQL è un database relazionale e utilizzando le chiavi esterne risulta semplice ottenere tutti i dati.
La scelta di questo approccio è motivata dal fatto che oggi, i database sono molto veloci e conviene eseguire meno query ma che richiedono una mole maggiore di dati dati, piuttosto che eseguirne di più, ma molto specifiche.

Il database è fondamentale in un applicativo come questo, perché, vista la mole di dati che vengono trattati, è necessario salvarli in ogni loro stato per garantirne la persistenza e la coerenza.

[4] https://www.postgresql.org/docs/
[5] https://orm.drizzle.team/docs/overview

4.2 Funzionalità Chiave

4.2.1 Autenticazione

L'autenticazione è una funzionalità fondamentale dell'applicativo, che permette di verificare l'identità dell'utente che sta effettuando la richiesta e di garantire che solo gli utenti autorizzati possano accedere ai servizi offerti dall'applicativo stesso.
La sessione è stata implementata utilizzando un plugin di Fastify chiamato Fastify-Session[6], che permette di gestirla utilizzando i cookies. Ogni volta che un utente effettua il login, viene generato un ID di sessione che viene salvato sul database, associandolo all'utente con cui si è loggato, e a cui viene dato un tempo di vita di una settimana. Questo permette all'utente di rimanere loggato anche dopo aver chiuso il browser e di non dover effettuare il login ad ogni accesso.

4.2.2 Caricamento dei Prodotti

Il processo di caricamento dei prodotti è gestito tramite un endpoint dedicato, a cui è possibile inviare i dati solamente se si inserisce nell'header della richiesta HTTP una chiave API valida. Questo permette che solo gli utenti autorizzati possano caricare i prodotti e che ogni richiesta sia tracciata e verificata.
I dati devono essere invati in formato JSON e situati all'interno del body della richiesta e una volta ricevuti, vengono caricati sul database, con lo stato iniziale di "Pending". Utilizzando la chiave API, tutti i prodotti caricati vengono associati all'Organizzazione a cui appartiene l'utente che ha effettuato la richiesta, in modo tale da avere la certezza che ogni utente possa vedere solo i prodotti caricati dalla propria Organizzazione.

4.2.3 Valutazione del Prodotto

Dopo che un prodotto è stato caricato e salvato sul database, si elaborano singolarmente i 3 Jobs derivati e si inviano all'IA per essere valutati. Questo processo è gestito dallo script ProductValidator che ogni 1250 millisecondi, controlla se ci sono dei Jobs in attesa di essere valutati e, se presenti, li invia all'IA utilizzando un SDK creato appositamente dal servizio esterno che permette di interfacciarsi in modo semplice con le intelligenze artificiali, specificando solamente:

- **Il modello da utilizzare**
- **Il prompt da inviare**

[6]https://www.npmjs.com/package/@fastify/session

- **Il contenuto da valutare**

Ogni risultato viene salvato sul database, associandolo al Job che lo ha generato, indicando il TraceID della valutazione fatta dall'IA, che identifica la richiesta singola e che permette di fare fine tuning all'IA. Di seguito un esempio di un Job inviato per essere valutato e della risposta data dal modello di IA (Fig 4.1)

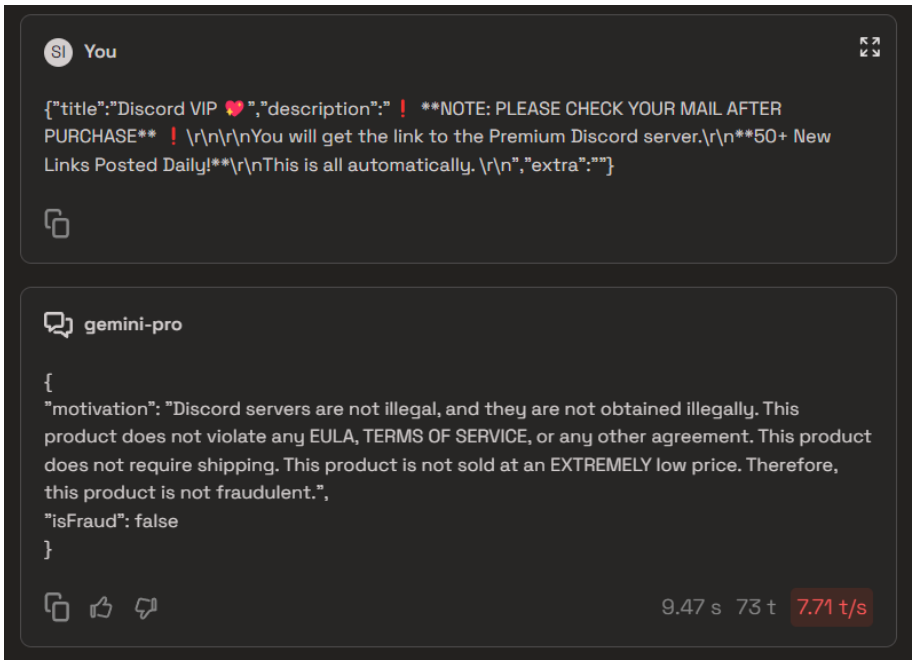

Figura 4.1: Esempio di schema di un Job inviato all'IA e della risposta ottenuta

4.2.4 Validazione Manuale

La validazione manuale nasce con lo scopo di addestrare il modello e migliorarne l'efficienza, in modo tale che possa lavorare completamente in autonomia e, quindi, non ci sia più necessità di controllare manualmente i prodotti valutati.
Il processo di validazione genera un webhook, se non già esistente e non ancora inviato, e un data point per ogni Job associato al prodotto in questione. Entrambi vengono caricati nelle rispettive tabelle del database, così da poter essere visti dagli script che li devono elaborare.
Come detto in precedenza, ogni Organizzazione può personalizzare la soglia di accuratezza minima per validare automaticamente il risultato dell'IA, senza l'intervento dell'utente, che però può sempre rigettare il risultato proposto.

4.2.5 Data Points

Un data point rappresenta un valore booleano che indica se il prodotto è stato valutato correttamente dall'IA, oppure se questa ha commesso un errore. Viene inviato utilizzando l'SDK di PremAI specificando il TraceID, che è un valore unico per ogni valutazione di un Job, e il valore booleano che rappresenta il risultato della valutazione.
I data point possono essere modificati e aggiornati in qualsiasi momento semplicemente creando un data point con un TraceID già esistente. Questo andrà a modificare quello creato in precedenza, senza crearne uno nuovo, andando in conflitto con quello già presente.

L'invio dei data point può fallire per vari motivi ed è quindi stato necessario implementare un sistema di retry che, in caso di errore, riprova ad inviare il datapoint dopo un periodo di tempo che aumenta esponenzialmente ad ogni tentativo, fino ad attendere al massimo una settimana prima di considerare il datapoint come perso.

Il modello di IA, si adatta in tempo reale ai feedback ricevuti, quindi dopo ogni data point creato o modificato acquisisce nuove informazioni e le risposte che darà da quel momento in poi saranno più accurate.

```
try {
  const res = await client.feedbacks.create({
    trace_id: dataPoint.trace,
    feedback: {positive: dataPoint.positive,}
  })
  if (res) {
    await dataPointService.update(dataPoint.id, {
      status: DATAPOINT_STATUS.COMPLETED,
    })
  } else {
    throw new Error("No trace id")
  }
}
catch (e) {
  console.log(e)
  if (dataPoint.try < ExponentialTimeHandler.maxTries) {
    await dataPointService.updateStatus(dataPoint.id, DATAPOINT_STATUS.ERROR)
  }
  else {
    await dataPointService.updateStatus(dataPoint.id, DATAPOINT_STATUS.FAILED)
  }
}
else {
if (env.ENVIROMENT === "DEVELOPMENT") {
  console.log("No data point")
}
```

Figura 4.2: Parte di codice utilizzata per la gestione dell'invio del data point, con aggiornamento dello stato in base all'esito

4.2.6 Invio dei risultati

La gestione dei webhooks per la restituzione dei risultati dei singoli prodotti è stata implementata per garantire all'Organizzazione la ricezione dei risultati delle valutazioni in modo tempestivo e affidabile.
Ogni webhook creato, viene caricato sul database, inizialmente in uno stato di "Pending", ad indicare che il webhook è pronto per essere inviato. Finché non viene restituito all'Organizzazione, può essere modificato dall'utente cambiando il valore della validazione del prodotto a cui fa riferimento.
Una volta inviato con esito positivo, il webhook assume uno stato definitivo, non più modificabile e, nel caso, ne verrà creato uno nuovo associato sempre allo stesso prodotto e sarà compito dell'Organizzazione gestire queste situazioni (relazione 1 prodotto ad n webhooks).

Come per i data points, anche l'invio dei webhooks può fallire per vari motivi e quindi, anche in questo caso, è stato implementato un sistema di retry che, in caso di errore, riprova ad inviare il webhook dopo un intervallo che aumenta esponenzialmente ad ogni tentativo, fino a raggiungere il limite massimo di una settimana prima di considerare la restituzione del risultato del prodotto come fallita e di conseguenza anche la valutazione complessiva del prodotto.

Capitolo 5

Valutazione

L'applicativo è in produzione dal 01/06/2024 e al momento gestisce una sola Organizzazione.
Nei mesi precedenti all'adozione, sono stati eseguiti test comparativi su differenti modelli di intelligenze artificiali e per ognuno di questi, sono stati valutati 5780 prodotti.
Per la valutazione del livello di accuratezza del modello di IA è stato utilizzato il seguente criterio di valutazione:

- Esclusi i modelli che non prevedevano la funzionalità di addestramento (fine tuning).
- Ove presente la possibilità di addestrare il modello, sono stati presi in considerazione gli esiti della valutazione dei primi 500 prodotti caricati e degli ultimi 500.

Questo metodo di valutazione ha permesso di identificare precisamente le differenze di comportamento dell'IA prima e dopo il processo di fine tuning, visto che i primi prodotti vengono valutati sul modello non ancora addestrato, mentre gli ultimi dopo che l'IA ha ricevuto più di 5000 feedback inerenti ai prodotti precedentemente valutati.

5.1 Confronto tra Intelligenze Artificiali

Il grafico in Fig. 5.1 mostra il confronto tra alcuni dei vari modelli utilizzabili.

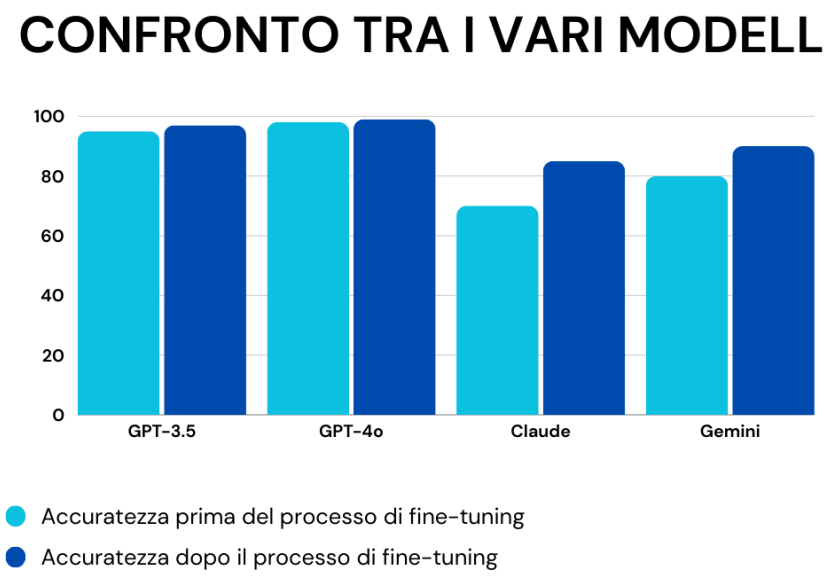

Figura 5.1: Grafico di confronto dei 4 modelli di Intelligenza Artificiale testati per la valutazione dei prodotti, che indica l'accuratezza prima e dopo il processo di fine-tuning

5.1.1 GPT

Utilizzando il modello GPT-3.5-turbo, con il modello non allenato si è ottenuta un'accuratezza del 95%, mentre, a seguito del fine-tuning si è raggiunta un valore del 97%.
Differisce leggermente da questi risultati, il modello GPT-4o che ha raggiunto sin da subito un'accuratezza del 98% sui primi prodotti valutati e che, grazie al processo di fine-tuning, ha raggiunto il 99% sugli ultimi 500. Questi risultati hanno indicato come entrambi i modelli si siano comportati in maniera efficace, ma abbiano sofferto di alcuni limiti. A seguito, infatti, di un'analisi, si è notato come la maggioranza dei prodotti, valutati in modo errato dal modello, fossero dei falsi positivi (quindi, prodotti illeciti o fraudolenti che però, l'intelligenza artificiale ha valutato come leciti, anche se con una percentuale di affidabilità raramente intorno al 100%). Il loro numero è diminuito a seguito del processo di fine tuning, ma meno delle aspettative.

5.1.2 Claude

Sono stati eseguiti dei test anche con il modello Claude-3-opus, che si è comportato decisamente peggio rispetto ai modelli GPT, infatti sui primi 500 prodotti valutati,

ha raggiunto un'accuratezza del 70% circa, mentre, a seguito del fine-tuning, non ha superato l'85%. Questo modello non è sempre riuscito a rispettare i vincoli e le regole imposte nel prompt, e questo ha portato sia a dei falsi negativi, sia a dei falsi positivi, che purtroppo non sono stati completamente risolti a seguito del fine-tuning.

5.1.3 Gemini

Il modello Gemini-pro, è stato una via di mezzo tra i modelli GPT e Claude, infatti, se sui primi prodotti ha avuto delle difficolta, raggiungendo un'accuratezza di circa l'80% si è comportato decisamente meglio a seguito del fine-tuning, raggiungendo un'accuratezza leggermente superiore al 90%. Anche in questo caso, i falsi positivi hanno caratterizzato la maggior parte degli errori, ma il modello si è adattato ai feedback migliorando sensibilmente le proprie prestazioni.

5.2 Considerazioni comuni a tutti i modelli

A seguito delle analisi condotte sui singoli modelli si possono trarre le seguenti conclusioni.
Se da un lato, è chiaro che i modelli GPT si sono comportati decisamente meglio rispetto a Claude e Gemini, dall'altro, è altrettanto chiaro che tutti e tre i modelli abbiano sofferto di problemi simili, come i falsi positivi, che non sono stati completamente risolti con il fine tuning.
I risultati però dimostrano come l'utilizzo di modelli basati su IA, sia decisamente migliore rispetto ai metodi tradizionali, come un approccio *rule based*, più semplice da aggirare, mettendo in atto delle strategie specifiche, come per esempio sostituire le lettere con dei numeri o con dei simboli, che possono sfuggire a un controllo statico, ma che sono più facilmente riconoscibili da un'intelligenza artificiale che in questo caso si è comportata molto bene e ha sempre interpretato correttamente il significato del testo.

Capitolo 6

Conclusioni

Il ricorso a modelli di intelligenza artificiale, in particolare a quelli dotati di capacità di autoapprendimento (fine tuning), si è dimostrato essere una strategia estremamente efficace per il riconoscimento di prodotti fraudolenti in siti di e-commerce.
L'implementazione di un'applicativo basato su queste tecnologie e integrato con i sistemi dei marketplace, costituisce una soluzione concreta e a basso costo per ottenere una significativa riduzione dei rischi di frode, con conseguente miglioramento del rapporto fiduciario tra marketplace e clienti.

La capacità di autoapprendimento dei modelli di IA sicuramente fornirà risultati sempre più accurati e affidabili tanto più verranno utilizzati. Anche l'evoluzione tecnologica e gli enormi investimenti che caratterizzeranno il settore nei prossimi anni, contribuiranno a rendere sempre più efficaci queste soluzioni. Sarà necessario, però, continuare a monitorare e presidiare lo sviluppo del settore che, come tutti i settori in rapida evoluzione, potrebbe presentare soluzioni di intelligenza artificiale sempre più sofisticate e performanti, rispetto a quelle che oggi sono risultate essere le migliori.

L'applicativo web trattato in questa tesi, oltre ad essere già diventato un prodotto di mercato che ha riportato risultati concreti, per rimanere competitivo dovrà necessariamente continuare a evolversi e ad adattarsi alle nuove tecnologie con una rapidità e una flessibilità che dovranno necessariamente essere superiori alle soluzioni tradizionali alle quali si era abituati.

In futuro, si potrebbe pensare di integrare l'applicativo con altri servizi esterni che permetterebbero di offrire nuovi modelli di IA, piuttosto che creare una pagina di dettaglio dei prodotti con informazioni più mirate e specifiche riguardanti i prodotti stessi, aggiungendo tutte le varie valutazioni effettuate dall'IA.
Sicuramente sarebbe interessante anche sviluppare un'applicazione mobile, che per-

metterebbe agli utenti di accedere ai servizi offerti dal marketplace in qualsiasi momento e da qualsiasi luogo, rendendo l'applicativo ancora più accessibile e fruibile.

Ringraziamenti

Ringrazio il Prof. Angelo Di Iorio per avermi guidato durante la stesura di questa tesi, dalla sua fase inIziale sino al suo completamento. Per aver mostrato sincero interesse all'argomento trattato e per essere stato sempre disponibile, anche di persona, fornendomi importanti consigli e suggerimenti in una fase così cruciale del mio percorso accademico.

Un sentito ringraziamento al correlatore della tesi, Giovanni Bruno, per la disponibilità mostrata e per il supporto dato durante la realizzazione del progetto e la stesura della tesi.

Un grazie anche all'azienda Soluzioni Futura Srl per avermi permesso di partecipare alla creazione di un progetto innovativo e interessante al fianco di un team fantastico, che mi ha aiutato a crescere e a migliorare, sia dal punto di vista professionale che personale.

Grazie alla mia famiglia, per avermi aiutato durante questo percorso, per avermi incoraggiato nei momenti più difficili e per avermi permesso di fare ciò che ritenevo più giusto, supportando ogni mia scelta.

Un grazie in particolare a mio fratello, che mi è stato d'ispirazione in questi anni e che, nonostante la distanza, è sempre stato presente.

Grazie alla mia ragazza che è sempre stata al mio fianco, che si è adattata alla situazione quando ne avevo bisogno e che mi ha supportato in tutto quello che ho fatto in questi 3 anni.

Infine grazie ai miei compagni di studio che, chi più recentemente e chi fin dall'inizio, sono stati una parte fondamentale del mio percorso di studi.

Capitolo 7

Riferimenti bibliografici

[1] A. J. Adetayo, M. O. Aborisade, and B. A. Sanni. Microsoft copilot and anthropic claude ai in education and library service. *Library Hi Tech News*, 2024.

[2] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), jul 2009.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[4] D. Coppola. E-commerce as percentage of total retail sales worldwide from 2021 to 2027. `https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/`, February 27, 2024. [Accessed 26-06-2024].

[5] J. Dai, H. Liu, and Q. Zhang. One class support vector machine active learning method for unbalanced data. In *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 309–315, 2020.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] V. N. Dornadula and S. Geetha. Credit card fraud detection using machine learning algorithms. *Procedia Computer Science*, 165:631–641, 2019. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019.

[8] A. Graves. *Long Short-Term Memory*, pages 37–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[9] P. Hajek, L. Hikkerova, and J.-M. Sahut. Fake review detection in e-commerce platforms using aspect-based sentiment analysis. *Journal of Business Research*, 167:114143, 2023.

[10] S. Handoyo. Purchasing in the digital age: A meta-analytical perspective on trust, risk, security, and e-wom in e-commerce. *Heliyon*, 10(8):e29714, 2024.

[11] R. Kiros, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler. Skip-thought vectors. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[12] I. Lauriola, A. Lavelli, and F. Aiolli. An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing*, 470:443–456, 2022.

[13] X. Li and L. Chen. Fake review detection using deep neural networks with multimodal feature fusion method. In *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 2869–2872, 2023.

[14] L. R. Medsker, L. Jain, et al. Recurrent neural networks. *Design and Applications*, 5(64-67):2, 2001.

[15] A. Patcha and J.-M. Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12):3448–3470, 2007.

[16] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[17] Preeti. Review on text mining: Techniques, applications and issues. In *2021 10th International Conference on System Modeling Advancement in Research Trends (SMART)*, pages 474–478, 2021.

[18] I. SADGALI, N. SAEL, and F. BENABBOU. Fraud detection in credit card transaction using machine learning techniques. In *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, pages 1–4, 2019.

[19] P. K. Sadineni. Detection of fraudulent transactions in credit card using machine learning algorithms. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 659–660, 2020.

[20] H. R. Saeidnia. Welcome to the gemini era: Google deepmind and the information industry. *Library Hi Tech News*, (ahead-of-print), 2023.

[21] Y. Santur. Sentiment analysis based on gated recurrent unit. In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, pages 1–5, 2019.

[22] F. Stahlberg. Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69:343–418, 2020.

[23] M. Sukanya and S. Biruntha. Techniques on text mining. In *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pages 269–271, 2012.

[24] J. Yao and M. Shepperd. Assessing software defection prediction performance: why using the matthews correlation coefficient matters. In *Proceedings of the 24th International Conference on Evaluation and Assessment in Software Engineering*, EASE '20, page 120–129, New York, NY, USA, 2020. Association for Computing Machinery.

[25] G. Yenduri, M. Ramalingam, G. C. Selvi, Y. Supriya, G. Srivastava, P. K. R. Maddikunta, G. D. Raj, R. H. Jhaveri, B. Prabadevi, W. Wang, A. V. Vasilakos, and T. R. Gadekallu. Gpt (generative pre-trained transformer)— a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 12:54608–54649, 2024.