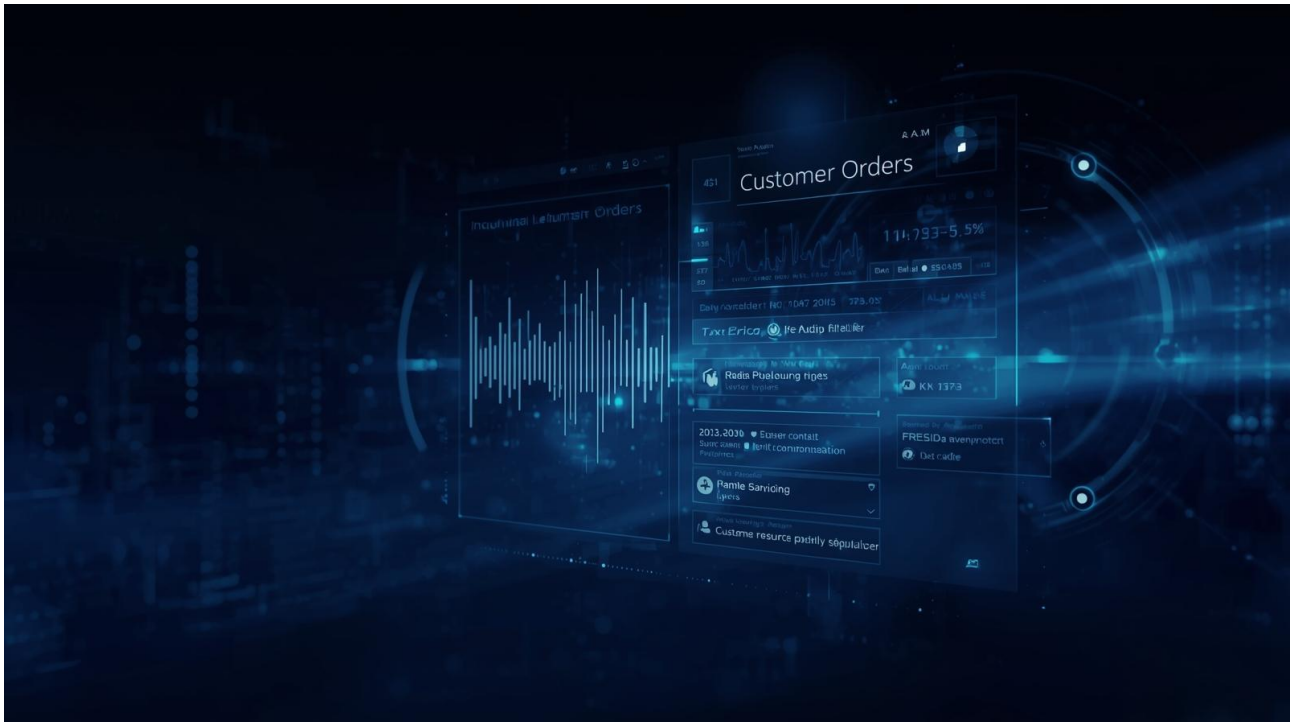




Università degli studi di Padova
Corso di Ingegneria del software
Prof. Tullio Vardanega



**SmartOrder: Analisi multimodale per la
creazione automatica di ordini**
Capitolato per il progetto didattico 2025/2026





Indice

1. Oggetto dell'appalto	2
2. Architettura AI Multimodale per la trasformazione di dati non strutturati in strutturati	4
2.1 Architettura proposta del sistema.....	5
3. Tecnologie, suggerimenti e risorse utili	7
4. Documentazione	10
5. Supporto	10
6. L'azienda: Ergon Informatica Srl	11

1. Oggetto dell'appalto

Nel panorama odierno, caratterizzato da un flusso costante di informazioni e da una molteplicità di canali comunicativi, le aziende si trovano quotidianamente ad affrontare la complessità di gestire richieste provenienti dai clienti in forme sempre diverse. Queste richieste possono arrivare sotto forma di testi scritti



(email, chat, moduli), messaggi vocali (registrazioni, telefonate trascritte) o immagini (fotografie di prodotti, documenti, cataloghi). Sebbene la varietà di input renda più immediato il contatto tra cliente e azienda, comporta anche la necessità di strumenti capaci di interpretare e standardizzare tali informazioni.

Il progetto si colloca nell'ambito della gestione degli ordini di acquisto da parte dei clienti di un'azienda. Gli ordini di acquisto sono le richieste con cui un cliente domanda la fornitura di determinati prodotti, specificando in genere l'articolo e la quantità desiderata. Nella pratica, però, le informazioni fornite non sempre risultano complete o precise: il cliente può descrivere il prodotto solo in parte oppure utilizzare termini diversi da quelli presenti nei cataloghi aziendali.

L'obiettivo del progetto è quindi automatizzare la ricezione di questi ordini, interpretarli correttamente e trasformarli in ordini cliente strutturati e pronti per l'inserimento nel database aziendale. La criticità principale è che i dati raccolti sono spesso non strutturati: non seguono uno schema definito e possono contenere ambiguità, ridondanze o informazioni incomplete. La gestione manuale di tali richieste, oltre a richiedere tempo e risorse, aumenta il rischio di errori che possono compromettere la soddisfazione del cliente e l'efficienza complessiva dei processi aziendali.

Il progetto **SmartOrder** nasce per rispondere a questa esigenza, proponendosi come una piattaforma intelligente in grado di **analizzare input multimodali** – testo, audio e immagini – e convertirli automaticamente in ordini strutturati. Grazie all'integrazione di tecniche avanzate di **intelligenza artificiale (AI)**, **machine learning (ML)** e **natural language processing (NLP)**, il sistema è capace di riconoscere le intenzioni del cliente, estrarre le informazioni rilevanti e validarle in maniera coerente.

Una volta processati, i dati vengono trasformati in un ordine cliente completo e normalizzato, pronto per essere archiviato all'interno di un **database** per essere utilizzato dai sistemi ERP¹ aziendali. Questo approccio consente di ridurre drasticamente l'intervento umano nelle fasi ripetitive e a basso valore aggiunto, migliorando la produttività e permettendo al personale di concentrarsi su attività strategiche.

L'adozione di **SmartOrder** porta benefici tangibili sia all'azienda che al cliente finale. Da un lato, le imprese ottengono maggiore velocità e precisione nella gestione degli ordini, con conseguente riduzione dei costi operativi; dall'altro, i clienti sperimentano un'esperienza di acquisto più fluida, semplice e affidabile, percependo un miglioramento significativo nella qualità del servizio.

In sintesi, questo progetto intende dimostrare come le tecnologie di intelligenza artificiale possano essere applicate con successo a processi reali di business, trasformando un compito complesso e frammentato in un flusso lineare, automatizzato e scalabile.

¹ ERP è l'acronimo di Enterprise Resource Planning, che in italiano significa "pianificazione delle risorse aziendali". Si tratta di un tipo di software gestionale integrato che aiuta le aziende ad automatizzare e gestire le attività quotidiane, come la contabilità, le risorse umane, la produzione, le vendite e la supply chain, centralizzando i dati in un unico sistema.



2. Architettura AI Multimodale per la trasformazione di dati non strutturati in strutturati

Il problema della gestione automatizzata degli ordini di acquisto può essere affrontato efficacemente mediante metodologie di **Machine Learning avanzato**, con particolare riferimento all'impiego di **Large Language Models (LLM)** e architetture multimodali. Gli LLM rappresentano tecnologie di intelligenza artificiale in grado di comprendere, analizzare e interpretare il linguaggio naturale, anche in presenza di testi complessi, ambigui o incompleti. Questa capacità li rende particolarmente adatti al contesto degli ordini cliente, in cui le descrizioni dei prodotti possono risultare non standardizzate o parziali.

In un sistema avanzato di gestione ordini, gli LLM operano come **motore semantico**, riconducendo descrizioni testuali non strutturate a entità coerenti presenti negli archivi aziendali. Ad esempio, un input testuale come "10 scatole di pasta integrale" può essere automaticamente mappato su un codice prodotto specifico del database, riducendo il rischio di errori manuali.

Per gestire **dati multimodali**, il progetto può integrare diverse tecnologie:

- **Testo:** oltre agli LLM per il NLP (ad esempio modelli basati su Transformer come GPT o BERT), si possono utilizzare pipeline di pre-processing con tecniche di tokenizzazione, normalizzazione e embedding semantici.
- **Immagini:** grazie a modelli di visione artificiale (ad esempio CNN² o Vision Transformers), è possibile estrarre informazioni da etichette, codici a barre o documenti scansionati, eventualmente convertendole in testo tramite OCR avanzato (ad esempio Tesseract o modelli basati su deep learning).
- **Audio:** per input vocali (ad esempio ordini telefonici o audio Whatsapp), si possono utilizzare sistemi di **speech-to-text** basati su reti neurali ricorrenti o Transformer, capaci di trascrivere e normalizzare il parlato in testo interpretabile dal modello.

Gli LLM e le reti multimodali possono essere addestrati sia con **apprendimento supervisionato** (per mapping diretto tra descrizioni e codici prodotto) sia con **apprendimento non supervisionato o auto-supervisionato** (per apprendere schemi linguistici e correlazioni tra diverse modalità di dati). L'uso di **embedding multimodali** consente di rappresentare testo, immagini e audio in uno spazio semantico condiviso, facilitando il matching tra input utente e catalogo prodotti.

Infine, il flusso operativo prevede che le informazioni estratte e interpretate dal modello vengano **normalizzate e strutturate** in un formato compatibile con il database gestionale (ad esempio JSON, XML o direttamente tabelle relazionali), consentendo l'inserimento automatico di ordini cliente validi, completi di quantità, codici prodotto e metadati associati (ad esempio modalità di inserimento, tipo di input, ecc.). Viene proposto in Figura 1 la struttura base del sistema.

² Una **CNN (Convolutional Neural Network)** è una rete neurale che estrae automaticamente le caratteristiche da un'immagine, permettendo di convertirla in rappresentazioni numeriche da cui altri modelli possono generare testo (ad esempio descrizioni o trascrizioni tramite OCR).



2.1 Architettura proposta del sistema

Il sistema è progettato come **pipeline modulare**, in grado di acquisire, elaborare e strutturare dati multimodali (testo, immagini, audio) provenienti da diversi canali, trasformandoli in ordini coerenti e pronti per l'inserimento in database ERP o gestionali. La progettazione prevede l'integrazione di **Large Language Models (LLM)**, modelli di visione artificiale e sistemi speech-to-text, con un focus sulla standardizzazione, validazione e arricchimento dei dati. Si propongono i seguenti layer che costituiscono il sistema (vedi Figura 2):

1. Layer di Raccolta Input

Il primo layer si occupa di acquisire i dati da canali diversi. Per il testo, il sistema raccoglie email, chat e moduli web; per le immagini, etichette prodotto, documenti scansionati e fotografie di ordini cartacei; per l'audio, chiamate telefoniche o messaggi vocali. Le tecnologie utilizzate includono sistemi di **OCR** (come Tesseract, EasyOCR o modelli basati su deep learning) per l'estrazione di testo dalle immagini e soluzioni di **Speech-to-Text** (ad esempio Whisper o Google Speech-to-Text) per la trascrizione dell'audio.

2. Layer di Pre-processing

I dati raccolti vengono normalizzati e puliti per ridurre rumore e ambiguità. Nel caso del testo, si effettuano tokenizzazione, lowercasing, rimozione di simboli non necessari e riconoscimento delle entità (NER³) per quantità, unità di misura e prodotti. Le immagini vengono ritagliate nelle aree informative, elaborate con OCR e corrette da eventuali distorsioni. L'audio viene trascritto accuratamente, con rimozione del rumore e normalizzazione linguistica.

3. Layer di Estrazione Feature / Embedding

In questa fase, i dati vengono trasformati in rappresentazioni vettoriali comprensibili dai modelli AI. Per il testo si generano embedding semantici tramite modelli LLM come BERT, RoBERTa o GPT. Le immagini vengono elaborate tramite CNN o Vision Transformer (ViT) per estrarre feature rilevanti, mentre l'audio può essere trasformato in embedding dal testo trascritto o direttamente da rappresentazioni audio.

4. Fusione Multimodale

Gli embedding di testo, immagini e audio vengono combinati in una rappresentazione vettoriale unica, tramite concatenazione o architetture di multimodal transformer con attention cross-modale. Questo embedding integrato permette un'interpretazione semantica coerente dei dati provenienti da più canali.

5. Modulo di Interpretazione Semantica

Il sistema utilizza l'embedded multimodale per mappare le informazioni sulle entità del catalogo aziendale. Vengono identificati i codici prodotto corretti, completate descrizioni incomplete o

³ Il **NER (Named Entity Recognition)** è una tecnica di **elaborazione del linguaggio naturale (NLP)** che identifica e classifica automaticamente nel testo le **entità nominate**, esempio: "Ordino 15 scatole di pasta Barilla da consegnare a Bologna entro il 20 settembre." **Output NER (entità estratte):**

- **Quantità: 15**
- **Unità di misura: scatole**
- **Prodotto: pasta Barilla**
- **Luogo di consegna: Bologna**
- **Data: 20 settembre**

ERGO N

ambigue e applicate regole aziendali e controlli preliminari. Questa fase si avvale di LLM fine-tuned su dati aziendali e logiche di validazione integrate.

6. Validazione e Arricchimento Dati

Prima dell'inserimento nei sistemi gestionali, i dati vengono verificati per garantire integrità, coerenza e completezza.

7. Output e Integrazione Database

I dati validati vengono trasformati in ordini strutturati pronti per i sistemi gestionali, in formati come JSON, XML o tabelle relazionali compatibili con ERP. L'inserimento avviene tramite API automatizzate, con tracciamento completo dei metadati, del canale di input e del livello di confidenza.

8. Monitoraggio e Feedback

Infine, il sistema implementa meccanismi di logging e feedback continuo, consentendo il retraining periodico e l'aggiornamento degli embedding multimodali basati su dati corretti o feedback manuale. Le regole aziendali vengono aggiornate in base ai pattern osservati, migliorando progressivamente accuratezza e affidabilità del sistema.

Il sistema è progettato con un'architettura che ne garantisce flessibilità, efficienza e affidabilità. Grazie alla **modularità**, ciascun layer può essere aggiornato o sostituito senza influire sull'intero sistema, permettendo integrazioni e migliorie rapide. La **scalabilità** consente di gestire grandi volumi di ordini e dati multimodali mantenendo elevate prestazioni. L'**affidabilità** è assicurata attraverso processi di validazione e arricchimento dei dati, che minimizzano errori e mantengono la coerenza delle informazioni. Infine, la **manutenibilità** è favorita da log dettagliati e meccanismi di feedback, che permettono interventi e ottimizzazioni continue senza interrompere il servizio.

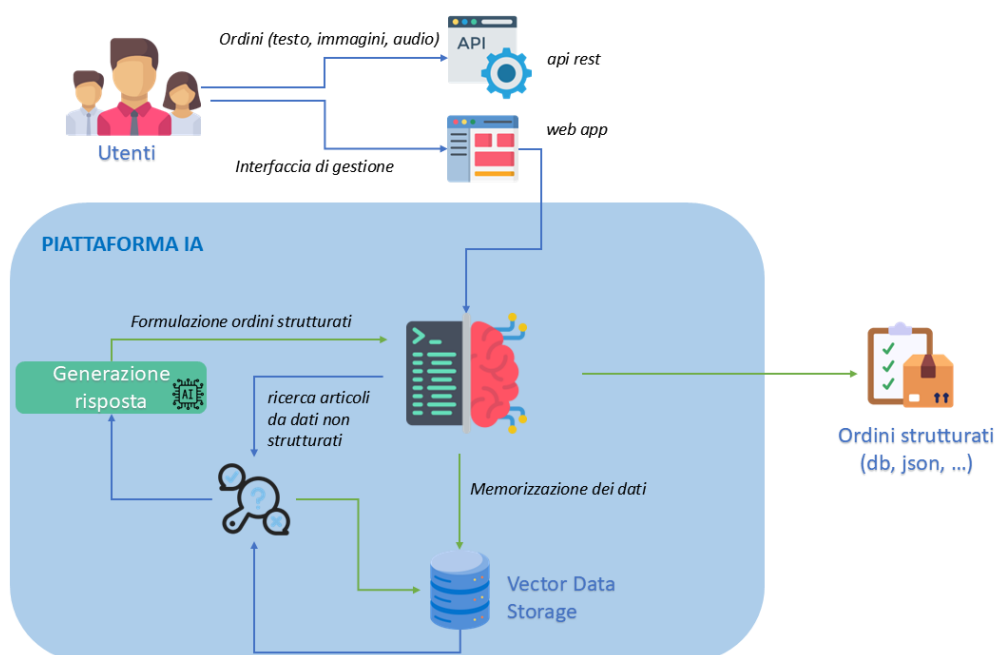


Figura 1 - Struttura base della piattaforma

ERGO N

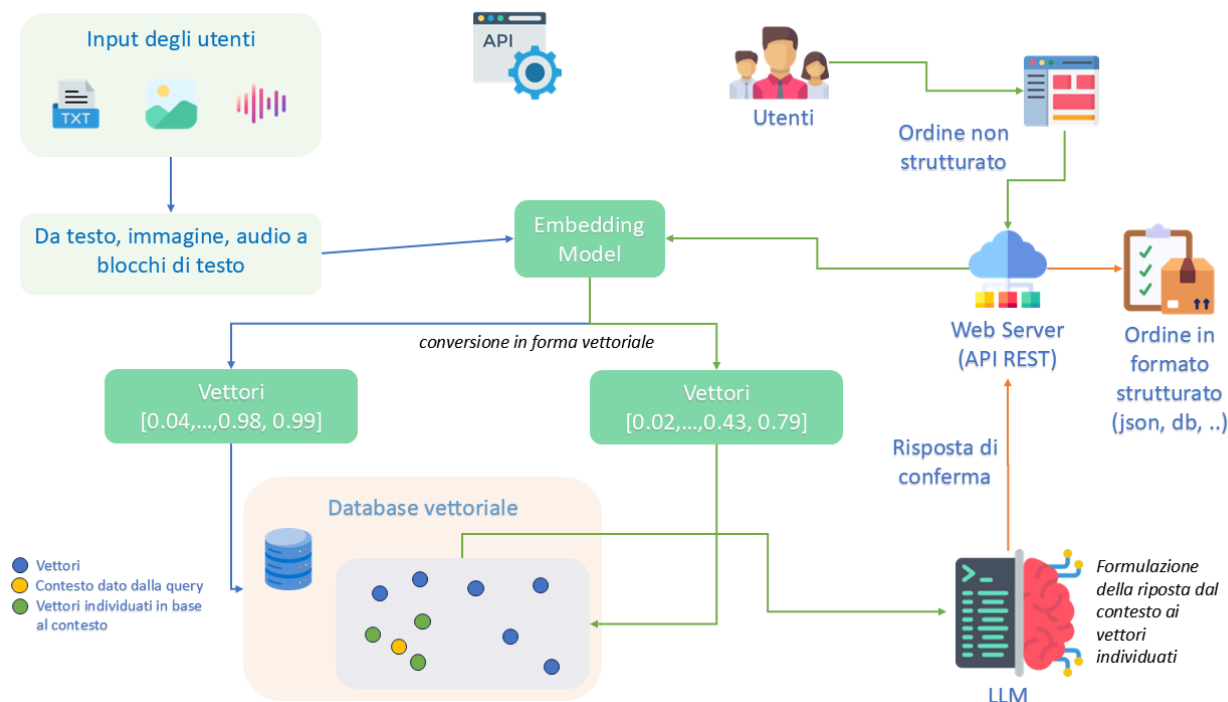


Figura 2 - Architettura del sistema

3. Tecnologie, suggerimenti e risorse utili

A richiesta degli studenti, l'azienda proponente è disponibile a fornire i dati di un caso di studio da utilizzare per lo sviluppo del progetto. I dati potranno essere dati in ingresso al sistema così da eseguire la fase di testing e poi interagire con il sistema per valutarne le prestazioni sfruttando un caso reale.

Considerata l'ampiezza e le potenzialità di sviluppo del progetto, si consiglia di iniziare concentrandosi su una o due modalità di input (ad esempio testo e audio), per poi estendere gradualmente le altre modalità di inserimento man mano che il sistema prende forma.

Di seguito vengono suggerite alcune tecnologie utilizzabili per il sistema esposto, è possibile comunque adottare tecnologie diverse, si veda la Figura 3 le aree proposte dove utilizzarle.

- **Database relazionale**

alla base del progetto può essere scelto un qualsiasi database relazione tra i vari disponibili sul mercato, ad esempio Sql Server Express, MySQL o MariaDB

- **Modelli di Linguaggio e NLP**

- **BERT (Bidirectional Encoder Representations from Transformers)**

Un modello di linguaggio bidirezionale pre-addestrato da Google, utilizzato per comprendere il contesto delle parole in una frase.

https://huggingface.co/docs/transformers/en/model_doc/bert

- **RoBERTa (Robustly optimized BERT approach)**



Una variante di BERT ottimizzata da Facebook AI, che migliora le prestazioni di BERT con nuovi obiettivi di pre-addestramento.

https://huggingface.co/docs/transformers/en/model_doc/roberta

- **GPT (Generative Pre-trained Transformer)**

Una serie di modelli di linguaggio sviluppati da OpenAI, progettati per comprendere il contesto, generare contenuti e ragionare attraverso testo, immagini e altro.

<https://platform.openai.com/docs/models>

- **Visione Computazionale e OCR**

- **Tesseract OCR**

Un motore OCR open-source che supporta vari formati di immagine e offre diversi formati di output.

<https://tesseract-ocr.github.io/>

- **EasyOCR**

Un modulo Python per estrarre testo da immagini, supporta oltre 80 lingue e vari script di scrittura.

<https://www.jaided.ai/easyocr/documentation/>

- **Convolutional Neural Networks (CNN)**

Una tipologia di rete neurale feedforward che apprende caratteristiche tramite l'ottimizzazione di filtri.

<https://medium.com/thedeephub/convolutional-neural-networks-a-comprehensive-guide-5cc0b5eae175>

- **Vision Transformer (ViT)**

Un'architettura di rete neurale che applica i trasformatori alla visione artificiale, trattando le immagini come sequenze di patch.

<https://arxiv.org/abs/2010.11929>

- **Riconoscimento Vocale e Trascrizione**

- **Whisper (OpenAI)**

Un sistema di riconoscimento vocale automatico addestrato su 680.000 ore di dati multilingue e multitask.

<https://openai.com/it-IT/index/whisper/>

- **Google Speech-to-Text**



Servizio di Google Cloud che converte l'audio in trascrizioni di testo, integrabile facilmente nelle applicazioni.

<https://cloud.google.com/speech-to-text/docs>

- **API REST**

Per la comunicazione tra il modello LLM e l'app di interazione con l'utente

- **Comunicazione da/per il database**

In funzione del componente scelto per lo sviluppo del modello LLM, si possono optare per diverse strade per comunicare con il database:

- Attraverso connettori standard una fonte dati ODBC
- Rendere indipendente il modello LLM dal database implementando un middleware che gestisca la comunicazione tra i componenti (es. JSON)

- **Interfaccia utente**

Sviluppo di una webapp responsive che implementa l'interfaccia di iterazione tra il sistema e l'utente:

- **.NET Blazor**

Framework che permette di sviluppare interfacce web interattive usando C#, unificando frontend e backend in un unico stack.

<https://dotnet.microsoft.com/it-it/apps/aspnet/web-apps/blazor>

- **React.js**

Framework molto popolare per costruire interfacce web interattive e componenti riutilizzabili.

<https://reactjs.org/>

- **Angular**

Framework completo basato su TypeScript, ideale se vuoi un'architettura più strutturata e modulare, con strumenti integrati per routing, servizi e test.

<https://angular.io/>

ERGON

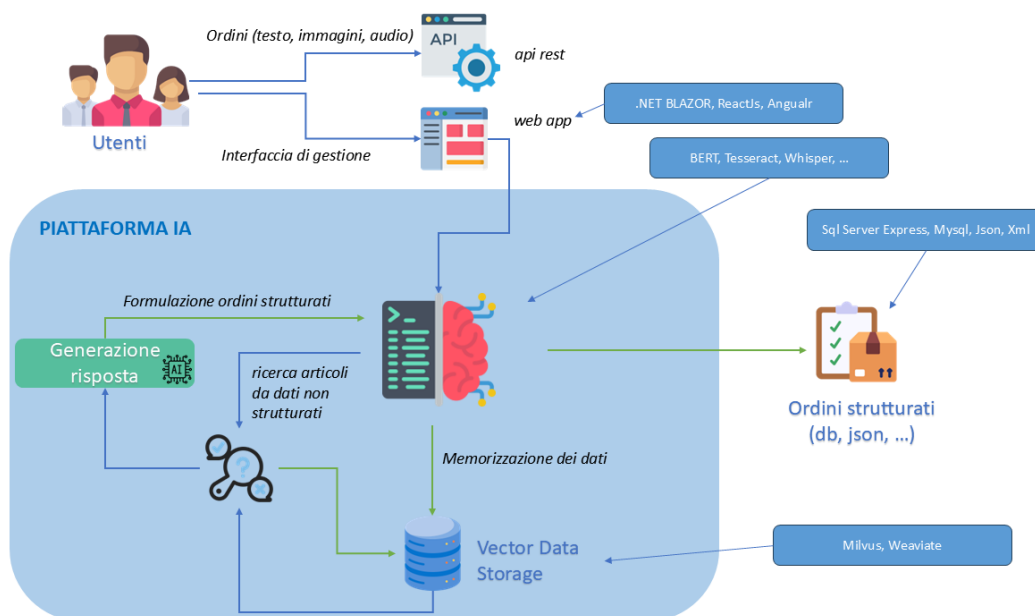


Figura 3 - Utilizzo tecnologie

4. Documentazione

Il progetto dovrà essere supportato dalla documentazione minima necessaria per comprendere il sistema sviluppato nelle sue componenti:

- Progettazione architetturale e tecnologie utilizzate
- Schema di progettazione della base di dati
- Codice prodotto in formato sorgente reso disponibile tramite repo pubblici (es. github)
- Documentazione descrittiva del sistema di raccomandazione implementato

5. Supporto

Per il progetto, l'azienda proponente fornirà ampio supporto da parte del team di R&D interno in varie fasi di sviluppo. L'interazione potrà avvenire sia nei locali aziendali sia da remoto tramite chat e/o chiamate. Agli studenti verrà assegnato un referente interno che li seguirà nelle varie fasi di sviluppo.

Contatti:

- Il referente del progetto sarà Gianluca Carlesso (carlesso@ergon.it).
- Per problemi di natura organizzativa: Anna Tieppo (tieppo.a@ergon.it)



6. L'azienda: Ergon Informatica Srl

Ergon Informatica Srl è una società di informatica di Castelfranco Veneto (TV), attiva dal 1988. Fornisce soluzioni ERP (Enterprise Resource Planner) per aziende del settore alimentare e di trasporto, con proprietari interamente analizzati e sviluppati al proprio interno. In particolare si occupa di progettazione, consulenza e sviluppo di software gestionale nelle seguenti aree: Amministrazione e Finanza, Controllo di Gestione, Area Vendite e Acquisti, Logistica, Produzione, Archiviazione Documentale, Qualità, Mobile, Web e Business Intelligence.

Ergon fornisce inoltre servizi sistemistici, installando tutta l'infrastruttura hardware necessaria ai propri clienti e occupandosi di sicurezza, networking, radio frequenza e virtualizzazione dei sistemi.

Negli ultimi dieci anni l'azienda è stata interessata da una costante crescita, sia in termini di Risorse Umane che di progetti, alcuni dei quali afferenti ad ambiti innovativi come l'AI (Artificial Intelligence) e il Machine Learning, nonché l'ottimizzazione dei sistemi gestionali. A questo proposito Ergon si avvale sempre di più della collaborazione di Università ed Istituti tecnici per offrire soluzioni software più competitive e all'avanguardia.

Si tratta di un'azienda certificata, che da molti anni segue gli standard ISO:9001, con un sistema di gestione e di processo secondo i parametri richiesti.