**ITS Digital Academy**  
**Relazione Tecnica**

**Progetto:** Progetto sentiment su Car Reviews



**Studente:** Guido Varnier  
**Data:** 19 giugno 2025  
**Docente/i di riferimento:** Giovanni Rossato  
**Sede:** Treviso

**1. Analisi EDA (Exploratory Data Analysis) e Preparazione dei dati**

La prima fase del progetto è stata la raccolta delle recensioni automobilistiche dal sito Carwow, tramite uno script Python con libreria BeautifulSoup. Sono state estratte 205 recensioni, contenenti:

* URL
* Titolo della recensione
* Prezzo
* Valutazione
* Testo

I dati sono stati salvati in formato CSV per facilitarne la pulizia e l’analisi.  
  
Successivamente, è stata effettuata un’attività di preprocessing con il modello Ollama, che ha eseguito l’analisi del sentiment in locale. Ogni recensione è stata suddivisa in 16 categorie specifiche, come: brand, modello, comfort, sicurezza, prezzo, affidabilità, ecc.  
  
Alcune risposte del modello non erano ben formattate, quindi è stato implementato un sistema di retry con prompt modificati, insieme a funzioni di pulizia basate su regex.

**2.** **Progettazione della Soluzione e dell’Architettura della Soluzione**

È stato realizzato un sistema modulare composto da:  
- Database relazionale (SQLite)  
- Pipeline di scraping e preprocessing  
- Modulo AI per la classificazione  
- Interfaccia utente con Streamlit  
  
Il database è stato normalizzato e progettato per garantire integrità dei dati e performance. La connessione è gestita da un modulo Python centralizzato, con configurazione tramite file `.env`.

**3.** **Progettazione UX/UI e Definizione dei KPI**

È stata sviluppata una dashboard interattiva tramite Streamlit, pensata per manager di brand automobilistici europei. Le funzionalità includono:  
- Filtri interattivi  
- Grafici comparativi tra brand europei e cinesi  
- Sezione KPI con metriche sintetiche  
  
Dal punto di vista estetico, sono stati usati colori distintivi (rosso per brand cinesi, verde teal per quelli europei) e layout multi-tab.

**4.** **Implementazione Tecnica**

Il progetto è stato sviluppato interamente in Python con le seguenti tecnologie:  
- BeautifulSoup per scraping  
- Pandas per manipolazione dati  
- Ollama per l’analisi del sentiment  
- Streamlit per la dashboard  
- Plotly per i grafici  
- SQLite per il database  
  
La dashboard è disponibile localmente.

**5.** **Testing e Validazione**

Il sistema ha mostrato buona efficacia, classificando correttamente oltre il 95% delle recensioni. Le dashboard risultano reattive con caricamenti inferiori a due secondi.  
  
Insight emersi:  
- I brand cinesi ottengono in media +0,3 punti rispetto a quelli europei  
- Trend positivo in crescita per i veicoli cinesi  
- Percezione migliore del rapporto qualità-prezzo per i marchi cinesi

**6. Problemi riscontrati e sviluppi futuri**

I problemi riscontrati includono:  
- Fonte unica (solo Carwow), con limitata rappresentatività  
- Alcuni output del modello AI mal formattati  
- Assenza di analisi temporale  
- Confronto limitato a un solo modello

Per migliorare il progetto si prevede di:  
- Estendere le fonti (AutoTrader, What Car)  
- Usare modelli ensemble più avanzati  
- Automatizzare scraping e aggiornamento live  
- Estendere l’analisi ad altri mercati europei