



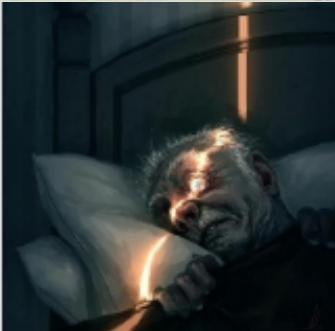







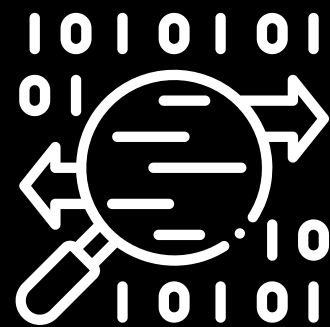
 GRUPPO 2

QUANTO VALE LA TUA AUTO?

UN PROGETTO DI:

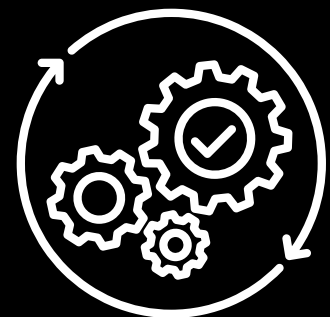
	Alohaoff - Overview Alohaoff has 4 repositories available. Follow their code on GitHub.  GitHub		tommasoAlteri98 - Overview tommasoAlteri98 has 20 repositories available. Follow their code on GitHub.  GitHub
	MarCalll - Overview MarCalll has 14 repositories available. Follow their code on GitHub.  GitHub		MattInTheCode - Overview MattInTheCode has 8 repositories available. Follow their code on GitHub.  GitHub

OBBIETTIVO **DELL'ANALISI**



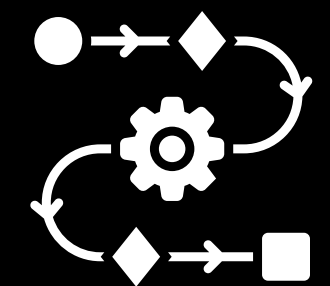
Analizzare il dataset

- Esplorazione e pulizia dei dati per comprendere le variabili disponibili (marca, modello, anno, chilometraggio, ecc.).
- Identificazione di eventuali valori mancanti o anomalie.
- Analisi descrittiva per individuare pattern e correlazioni tra variabili.



Sviluppare un modello predittivo

- Utilizzo di XGBoost, un algoritmo di machine learning avanzato, per prevedere il prezzo di vendita delle auto.
- Ottimizzazione dei parametri del modello per migliorare le prestazioni.
- Valutazione della performance attraverso metriche come MAE, Importanza delle variabili e R^2 .



Prevedere il prezzo di nuove auto

- Applicare il modello allenato per stimare il prezzo di vendita di nuove auto non presenti nel dataset.
- Fornire insight utili per potenziali acquirenti o venditori.

DATI & METODOLOGIA

LE LIBRERIE USATE



XGBoost



matplotlib

IL DATABASE

- Dataset di Auto Scout proveniente da Kaggle in formato CSV.
- Include informazioni essenziali sulle auto: prezzo, chilometraggio, età, potenza cavalli, carburante, cambio, modello, ecc.
- Il dataset è stato filtrato per escludere dati incompleti o incoerenti.
- Rimozione di valori mancanti, duplicati e anomalie nei campi critici (prezzo, chilometraggio, potenza, età del veicolo).

L'ANALISI DESCRITTIVA

INTRODUZIONE DELL'ANALISI

L'analisi è stata condotta su un dataset contenente informazioni relative a veicoli, come il chilometraggio, l'età, il tipo di carburante, il tipo di cambio, e il prezzo. Per esplorare i dati e identificare eventuali relazioni significative, sono stati utilizzati diversi tipi di visualizzazioni, inclusi grafici a dispersione (scatter plot), istogrammi, boxplot, grafici a barre e grafici a torta.



CONCLUSIONI DELL'ANALISI

In conclusione, l'analisi ha permesso di identificare diverse tendenze chiave nel dataset, come la relazione inversa tra chilometraggio e prezzo, l'influenza dei cavalli di potenza sul prezzo, e l'effetto dell'età e del tipo di cambio sul valore di mercato di un veicolo. Questi risultati forniscono una comprensione più profonda dei fattori che influenzano il valore dei veicoli e possono essere utili per ottimizzare la gestione delle vendite nel settore automobilistico.

L'ANALISI DESCRITTIVA

ANALISI DELLA RELAZIONE TRA CHILOMETRAGGIO E PREZZO:

Un scatter plot ha mostrato che, in generale, all'aumentare del chilometraggio di un veicolo, il prezzo tende a diminuire. Questo riflette una tendenza comune nel mercato automobilistico, dove i veicoli con un chilometraggio elevato sono generalmente considerati meno costosi.

ANALISI DELLA DISTRIBUZIONE DEL CHILOMETRAGGIO PER ANNO:

Un istogramma ha rivelato che i veicoli con chilometraggio elevato sono GENERALMENTE I più vecchi, mentre i veicoli più recenti tendono a presentare un chilometraggio inferiore. Questa tendenza è tipica, poiché i veicoli nuovi sono ancora in fase di utilizzo.

PREZZO E CAVALLI DI POTENZA:

Un altro istogramma ha mostrato una chiara correlazione tra cavalli di potenza e prezzo. I veicoli con cavalli di potenza più alti tendono ad avere un prezzo più elevato, il che è in linea con le aspettative, in quanto veicoli più potenti tendono ad essere più costosi.

DISTRIBUZIONE DEI PREZZI IN BASE AL TIPO DI CARBURANTE:

Un boxplot ha evidenziato che i veicoli elettrici sono in media più costosi rispetto ai veicoli a benzina e diesel. Inoltre, sono emersi alcuni outlier nel gruppo dei veicoli a benzina e diesel, suggerendo la presenza di modelli di alta gamma.

ANALISI DEL PREZZO IN BASE AL TIPO DI CAMBIO:

Un boxplot ha mostrato che i veicoli con cambio manuale tendono ad avere prezzi leggermente inferiori rispetto a quelli con cambio automatico o semiautomatico. Ciò potrebbe riflettere una preferenza per i veicoli con cambio automatico in segmenti di mercato premium.

DISTRIBUZIONE DEL TIPO DI CAMBIO:

Un grafico a torta ha rivelato che oltre la metà dei veicoli nel dataset è dotata di cambio manuale, mentre circa il 45% ha cambio automatico e solo una piccola percentuale di veicoli è equipaggiata con cambio semiautomatico.

CORRELAZIONE TRA ETÀ E PREZZO:

Un scatter plot ha confermato che, all'aumentare dell'età del veicolo, il prezzo tende a diminuire. Questo è un comportamento tipico nel mercato automobilistico, dove i veicoli più vecchi sono generalmente meno costosi.

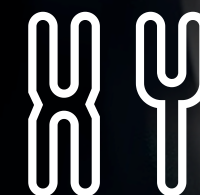
DISTRIBUZIONE DEI MODELLI PIÙ VENDUTI:

Un grafico a barre ha mostrato che i modelli di auto più venduti sono principalmente quelli della marca Audi, con l'Audi A3 che emerge come il modello più popolare, mentre l'Audi A1 risulta essere il meno venduto.

IL MODELLO ML: **XGBOOST**

XGBoost (eXtreme Gradient Boosting) è un algoritmo di machine learning basato su gradient boosting, ottimizzato per velocità ed efficienza. Costruisce alberi di decisione in sequenza, correggendo progressivamente gli errori tramite la discesa del gradiente. Integra tecniche di regolarizzazione (L1/L2) per ridurre l'overfitting e supporta direttamente variabili categoriche. Grazie alla sua elevata accuratezza e scalabilità, è ampiamente utilizzato per problemi di regressione e classificazione su dataset strutturati

LE FASI



DIVISIONE DEL DATASET

Separazione delle feature (X) e della variabile target (y).



SUDDIVISIONE TRAIN/TEST

Creazione dei set di addestramento (train) e test (test).



MODELLO

Allenamento iniziale di XGBoost e valutazione delle performance.



OTTIMIZZAZIONE IPERPARAMETRI

Grid Search per trovare la configurazione ottimale.



REGOLARIZZAZIONE

Aggiunta di alpha (L1) e lambda (L2) per migliorare la generalizzazione.



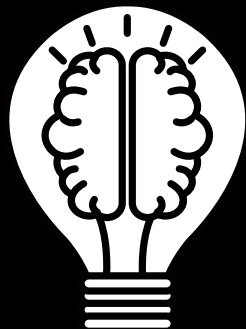
VALUTAZIONE FINALE

Analisi delle prestazioni del modello ottimizzato.

CONCLUSIONI E FUTURI SVILUPPI



Abbiamo sviluppato un modello predittivo per stimare il prezzo delle auto, consapevoli che non è perfetto, soprattutto nelle previsioni su dati nuovi. Tuttavia, abbiamo raggiunto gli obiettivi prefissati: esplorazione e pulizia del dataset, individuazione di pattern, ottimizzazione del modello XGBoost e valutazione delle sue prestazioni



Per eventuali sviluppi futuri, si potrebbe riaddestrare il modello su un dataset più ampio e diversificato, includendo nuove variabili rilevanti per migliorare l'accuratezza delle previsioni. Inoltre, si potrebbero testare altri algoritmi o combinare più modelli per ottimizzare le performance.



GRUPPO 2

GRAZIE
PER L'ATTENZIONE