

Car price predictor

Di Picello Davide e Fantin Anthony

Prefazione

Questo progetto nasce al di fuori di qualsiasi ambito scolastico/accademico, a partire da una necessità reale di Anthony, uno dei due ragazzi che ha sviluppato questa idea.

In quel periodo aveva iniziato ad avvicinarsi al mercato delle auto usate, in quanto voleva comprarsene una ma, ad ogni annuncio, era sempre in dubbio sull'effettiva correttezza del prezzo. Era giusto ? Era troppo ? Era troppo poco ?

L'unico modo per capirlo era cercare altri annunci simili e verificare se il prezzo fosse in linea.

Un giorno, però, gli venne un'idea che propose al suo amico Davide.

E se al posto di dover verificare ogni volta il prezzo, manualmente, avessero creato un software che lo calcolasse al posto suo ?

L'idea piacque ad entrambi e così iniziarono a svilupparla.

Sviluppo

L'idea

Anthony, studente di Diritto e Tecnologia, e Davide, studente di Informatica, entrambi all'Università di Padova, ma soprattutto grandi amici fin da quando erano bambini, iniziarono così a scambiarsi idee e punti di vista per la realizzazione di questa idea. I loro differenti background stimolarono ancora di più questo confronto.

Alla fine decisero di creare un software in grado di calcolare ogni volta il prezzo corretto di un'auto usata, in base ai seguenti parametri:

- Marca
- Modello
- Chilometri
- Anno di produzione

Si sarebbero potuti usare molti più parametri come il tipo di cambio, il tipo di carburante, la potenza del motore, il numero di porte, classe di emissioni, interni ...

Ma già solo con questi pochi parametri, il livello di accuratezza raggiunto era tale che si è deciso di tenere solo questi.

Realizzazione

Data scraping

Per rendere possibile tutto ciò è stato deciso, quindi, di creare una rete neurale che, in base al suo dataset di allenamento, avrebbe poi calcolato un prezzo.

Prima di tutto occorre, quindi, dotarsi di una buona base di dati.

Per farlo abbiamo creato un web scraper, ovvero un programma software, in grado di recuperare informazioni da pagine web. In particolare, il nostro software, è stato progettato per funzionare sul famosissimo sito di compravendita di auto usate “AutoScout24” che con il suo database di 2 milioni di auto in vendita in tutta europa (di cui 400 mila in Italia) era senza dubbio il luogo con più dati di nostro interesse che potessimo trovare sul web.

E dopo ore ed ore in cui i nostri computer continuavano a ricavare dati, arrivammo ad avere un database di 160 mila auto con relative caratteristiche e prezzo.

Un’ottima base da cui partire per addestrare il nostro modello

Modello

Il backend di tutto il progetto è stato scritto in Python 3.11, ottimo linguaggio per questo tipo di creazioni data la grande mole di librerie dedicate alla data science ed al data management.

Il modello in sé è una rete neurale addestrata con un algoritmo Backpropagation. Il numero di layer e di neuroni è sensibilmente cambiato con il passare delle fasi di testing, ma sono state utilizzate, in maniera pressoché fissa, tecniche quali: regularization, drop-out e funzioni di attivazione (es ReLU). Ciò che più di tutto ha richiesto impegno da parte del team è la parte di data processing e scelta di variabili e target.

La manipolazione dei dati è stata eseguita nella seguente modalità:

- scaling dei chilometri usando Robust Scaler
- scaling dell’anno usando Robust Scaler
- encoding tramite una tecnica chiamata One Hot Encoding delle variabili nome e modello della vettura
- (in fase di lavoro) encoding della variabile carburante

Il target è il prezzo della vettura sulla quale il modello è stato addestrato a riconoscere i pattern dell'auto: dai prezzi delle autovetture di lusso, dal deprezzamento delle utilitarie all'apprezzamento delle vetture più rare.

Nella fase finale il modello riconosce circa 2000 modelli di vetture con un'accuratezza vicina al 95%.

Rilascio

Al tutto è stata poi data un'interfaccia tramite una pagina web scritta in Flask disponibile al seguente [link](#).

Configurare un server è stata un'attività molto educativa che non capita di fare tutti i giorni.

Al momento la pagina web ha lo stretto necessario per funzionare, non sono state implementate alcune migliorie estetiche.

Conclusioni

Questo progetto è stato estremamente formativo e divertente, usare nuove tecnologie ed interfacciarsi ogni giorno con nuovi problemi di ogni tipo ci ha fatto comprendere quanto sia stimolante, ma anche impegnativo, portare un prodotto dalla nascita alla distribuzione.

Farlo, poi, con un amico, è stato ancora più bello e significativo.