

# Cross-Selling Assicurativo

**AssurePredict** è una compagnia di assicurazioni leader nel settore, specializzata nell'offrire soluzioni innovative per la gestione del rischio. Questo progetto mira a creare un modello predittivo in grado di individuare potenziali opportunità di **cross-selling** per clienti esistenti, identificando quelli che potrebbero essere interessati ad acquistare una polizza aggiuntiva per il loro veicolo.

## Obiettivo del Progetto

L'obiettivo è sviluppare un **modello di machine learning** che preveda se i clienti, che attualmente hanno un'assicurazione sanitaria, potrebbero essere interessati a sottoscrivere una polizza assicurativa per il loro veicolo. Il modello aiuterà **AssurePredict** a migliorare l'efficacia delle proprie strategie di cross-selling e ad aumentare la penetrazione nel mercato.

## Valore aggiunto per AssurePredict:

- **Aumento del tasso di conversione** nelle vendite di polizze auto.
- **Ottimizzazione delle campagne di marketing**, indirizzando le offerte a clienti più propensi ad acquistare.
- **Riduzione dei costi** legati a campagne di marketing inefficaci, grazie alla targettizzazione precisa.

## Dataset

Il dataset (scaricabile da qui: <https://proai-datasets.s3.eu-west-3.amazonaws.com/insurancecrosssell.csv>) contiene informazioni dettagliate sui clienti e sul loro comportamento assicurativo. Le caratteristiche principali del dataset sono:

- **id**: identificativo univoco del cliente.
- **Gender**: sesso del cliente.
- **Age**: età del cliente.
- **Driving\_License**: 1 se il cliente possiede la patente di guida, 0 altrimenti.
- **Region\_Code**: codice univoco della regione di residenza del cliente.
- **Previously\_Insured**: 1 se il cliente ha già un veicolo assicurato, 0 altrimenti.
- **Vehicle\_Age**: età del veicolo del cliente.
- **Vehicle\_Damage**: 1 se il cliente ha avuto incidenti o danni al veicolo in passato, 0 altrimenti.
- **Annual\_Premium**: importo annuale del premio assicurativo pagato dal cliente.
- **PolicySalesChannel**: canale utilizzato per la vendita della polizza (es. email, telefono, di persona).
- **Vintage**: giorni da cui il cliente è assicurato con **AssurePredict**.
- **Response**: 1 se il cliente ha accettato la proposta di cross-sell, 0 altrimenti.

## Attività Richieste

### 1. Esplorazione del Dataset

L'esplorazione preliminare del dataset permetterà di comprendere meglio la distribuzione delle caratteristiche e delle variabili target. In particolare, si analizzeranno:

- La **distribuzione della variabile "Response"**, per identificare eventuali sbilanciamenti tra clienti che accettano o rifiutano l'offerta di cross-sell.
- Le relazioni tra variabili chiave come **Annual Premium**, **Vehicle Age**, **Previously Insured**, e la risposta del cliente.

**Valore aggiunto:** Un'accurata esplorazione dei dati permette di identificare pattern nascosti e punti critici che influenzeranno il successo del modello predittivo.

### 2. Gestione dello Sbilanciamento delle Classi

La variabile target "Response" potrebbe essere sbilanciata, con molti più clienti che rifiutano l'offerta rispetto a quelli che la accettano. Per affrontare questo problema, verranno utilizzate tecniche di:

- **Class Weights**: penalizzazione della classe più frequente nel modello.
- **Oversampling o Undersampling**: creazione di un dataset più bilanciato per migliorare la capacità del modello di generalizzare.

**Valore aggiunto:** Gestire correttamente lo sbilanciamento delle classi è cruciale per evitare modelli che abbiano un alto tasso di falsi negativi, migliorando così la precisione del cross-sell.

### 3. Costruzione del Modello Predittivo

Utilizzando algoritmi di machine learning, verrà costruito un modello che predice la probabilità che un cliente risponda positivamente all'offerta di cross-sell.

**Valore aggiunto:** Il modello predittivo permetterà a **AssurePredict** di identificare con precisione i clienti più propensi a sottoscrivere una polizza aggiuntiva, migliorando così il ritorno sull'investimento delle campagne di marketing.

## Conclusione

Questo progetto permetterà a **AssurePredict** di sfruttare le potenzialità del **machine learning** per identificare opportunità di cross-selling in modo efficace e mirato. L'adozione di un approccio data-driven per la predizione delle risposte dei clienti garantirà non solo un aumento delle vendite, ma anche una maggiore soddisfazione del cliente grazie a offerte più pertinenti e personalizzate.

## Punti di attenzione

Fai attenzione alla distribuzione delle classi, dai uno sguardo a [questo approfondimento](#). In caso di classi sbilanciate puoi provare a:

- Penalizzare la classe più frequente (ricorda l'argomento `class_weight`)
- Utilizzare l'oversampling o l'undersampling (cfr. [questo articolo](#))

**Modalità di consegna:** Link pubblico a notebook di Google Colab