

Predizione successo per Campagna di Marketing

Il settore finanziario è estremamente concorrenziale e i piccoli risparmiatori sono spesso sottoposti a nuove offerte e promozioni da parte degli istituti di credito.

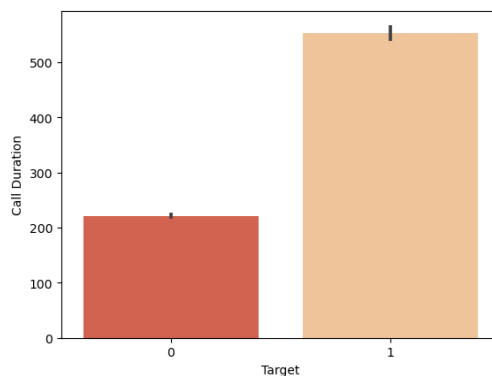
Il marketing diventa quindi essenziale, e una buona strategia integrata non può fare a meno di una campagna telefonica, utile sia ad instaurare un primo contatto diretto con un potenziale cliente, sia ad offrire nuove soluzioni ai propri clienti.

Occorre selezionare con cura le persone più propense a stipulare nuovi accordi con la banca sia per limitare i costi da sostenere per la campagna, che per evitare di contattare chi non è interessato ad un certo servizio. Evitare questi due inconvenienti potrebbe procurare dei vantaggi competitivi tangibili.

A tale scopo è possibile costruire un modello di regressione binaria capace di stimare la probabilità di successo della campagna di marketing verso un particolare individuo, conoscendo le sue caratteristiche. Una volta stimata la probabilità di successo, è possibile individuare i clienti più propensi alla sottoscrizione classificando chi è opportuno contattare e chi no. In questo modo, potrebbe aumentare notevolmente l'efficacia della campagna.

I dati disponibili [1] contengono 20 variabili (19 regressori e la variabile target) relative a 41.189 persone contattate. I regressori comprendono alcune informazioni personali di ogni contatto, della campagna oggetto di studio e dati macroeconomici registrati al momento della chiamata.

Per sfruttare l'informazione relativa alla durata della chiamata (Call Duration, esclusa dal modello predittivo [2]), si può notare dai dati disponibili che le persone che sottoscrivono il contratto in media rimangono in chiamata per 9 minuti, mentre chi non accetta la proposta in media termina la chiamata in meno di 4 minuti.



Abbiamo quindi implementato un modello di regressione binaria con scopo predittivo, focalizzandoci sugli individui che con maggior probabilità sottoscriveranno il prodotto finanziario.

Modello [4]:

Dall'analisi dei coefficienti, si nota un'influenza della situazione macroeconomica sull'efficacia della campagna: l'aumento dell'inflazione e della confidenza dei consumatori aumenta la tendenza a sottoscrivere il deposito, mentre l'aumento dell'indice di riferimento (Euribor a 3 mesi) ne diminuisce l'appetibilità e quindi il successo. Per massimizzare i risultati si suggerisce di avviare le future campagne in contesti di fiducia da parte dei piccoli risparmiatori.

Dalla regressione si nota un miglior risultato ottenibile dal contatto attraverso cellulare, piuttosto che per telefono fisso.

Inoltre, il modello evidenzia una maggiore potenzialità della campagna se effettuata nel mese di marzo, mentre mostra una situazione opposta nei mesi di maggio, agosto e settembre. Questo potrebbe dimostrare una minor propensione dei clienti nei mesi subito precedenti e successivi alle vacanze estive dove solitamente si investe una buona parte della disponibilità economica.

Per quanto riguarda la partecipazione alle campagne di marketing precedenti, i clienti che in passato hanno già rifiutato la proposta della campagna immediatamente precedente, generalmente rifiutano anche la successiva.

Si nota che alcune tipologie di lavoro hanno un'influenza negativa nell'effettivo successo della campagna. In particolare, per le categorie: casalinga, libero professionista e disoccupato. Le motivazioni potrebbero essere differenti, ipoteticamente un disoccupato potrebbe avere una disponibilità economica limitata in quel momento e non essere interessato ad un investimento.

Il libero professionista invece potrebbe essere più propenso al rischio mentre una casalinga probabilmente non ha una cultura finanziaria adeguata. Si potrebbe quindi consigliare di integrare alla campagna di marketing una campagna di comunicazione per far comprendere il potenziale dei prodotti finanziari offerti.

Esempio del modello in azione:

Laura lavora nel terziario, single, con casa di proprietà e un credito, contattata per la prima volta di mercoledì nel mese di maggio, in un periodo di scarsa fiducia dei risparmiatori. Il modello ha correttamente predetto la non sottoscrizione del deposito da parte della signora Laura, come effettivamente è accaduto.

Potenziali aggiustamenti in futuro:

Il nostro modello al momento è stato definito volutamente come "neutro", ovvero né sbilanciato verso una strategia di marketing estremamente espansiva, né troppo restrittiva.

In futuro e in base alle esigenze strategiche, sarà possibile modificare la soglia di classificazione e di conseguenza orientare i risultati prodotti dal modello, ad esempio spingendo verso una ricerca più circoscritta di persone potenzialmente interessate.

Variabile	Coefficiente
Intercetta	2.09
N. contatti durante la campagna	-0.19
Giorni dall'ultimo contatto	-0.05
Numero di contatti precedenti	0.12
Indice dei prezzi al consumo	0.27
Indice confidenza dei consumatori	0.28
Euribor 3 mesi	-0.86
Mutuo sulla casa	-0.08
Credito in essere	-0.25
Contatto telefonico	-0.81
Mese	
Marzo	0.82
Aprile	-0.09
Maggio	-0.80
Luglio	-0.23
Agosto	-0.77
Settembre	-0.57
Ottobre	0.17
Novembre	-0.53
Dicembre	-0.11
Giorno della settimana	
Lunedì	-0.54
Mercoledì	-0.21
Giovedì	-0.34
Partecipazione campagne prec.	
Campagna fallita	-1.90
Non partecipazione	-1.02
Stato Coniugale	
Divorziato	-0.46
Sposato	-0.20
Lavoro	
Operaio	-0.58
Imprenditore	-0.61
Casalinga	-0.81
Manager	-0.60
Pensionato	-0.15
Libero Professionista	-0.62
Servizi	-0.54
Studente	-0.34
Disoccupato	-0.78

Damiano Agachi Menna
Annalisa Basta
Simone De Bonis
Chiara Mercati
Daniele Torregiani

Appendice

[1] Trasformazione dei dati

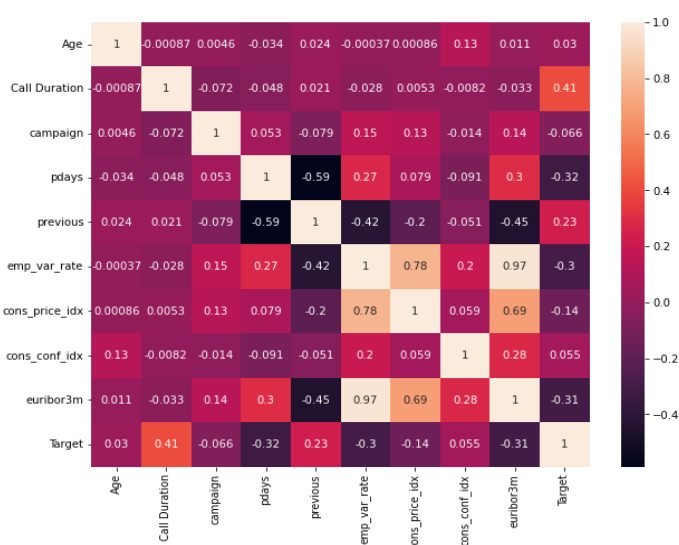
La variabile "Pdays" assegna '999' ai clienti che non sono già stati contattati da una precedente campagna; per questo è stata moltiplicata per una variabile di controllo binaria, creata in modo tale da azzerarne il valore nel caso in cui si manifesti il valore '999'.

Una volta eliminati i valori duplicati, le variabili quantitative sono state standardizzate.

[2] Informazioni Ridondanti

Abbiamo escluso il tasso di variazione dell'occupazione (emp.var.rate), in quanto altamente correlato con le variabili riferite all'indice dei prezzi al consumo e all'euribor (cons_price_idx e euribor3m).

Inoltre, abbiamo escluso dall'analisi anche la variabile relativa alla durata della chiamata (Call Duration), dal momento che l'informazione non è disponibile al momento della previsione.



[3] Sbilanciamento dei dati

La maggior parte delle persone contattate non ha sottoscritto il deposito (89% del totale).

Per risolvere lo sbilanciamento delle classi, abbiamo svolto un undersampling utilizzando l'algoritmo k-means con un numero di cluster pari a 8, dai quali abbiamo estratto il 25% delle istanze. Successivamente, come oversampling, abbiamo applicato l'algoritmo SMOTE considerando cinque vicini.

[4] Costruzione modello

Sul dataset bilanciato abbiamo eseguito una regressione logistica lasso per selezionare le variabili più importanti, in seguito abbiamo eseguito una logistica non penalizzata.

[5] Matrice di confusione e metriche

Veri/Previsti	0	1
0	4156	926
1	279	487

	Precision	Recall	F1
0	0.94	0.82	0.87
1	0.34	0.64	0.45
Accuracy			0.79
Macro avg	0.64	0.73	0.66
Weighted avg	0.86	0.79	0.82