

UNIVERSITA' DEGLI STUDI DI TORINO

DIPARTIMENTO DI CULTURE, POLITICA E SOCIETÀ
DIPARTIMENTO DI INFORMATICA
DIPARTIMENTO DI MATEMATICA "GIUSEPPE PEANO"
DIPARTIMENTO DI ECONOMIA E STATISTICA "COGNETTI DE MARTIIS"

MASTER DI I LIVELLO IN ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE ANNO ACCADEMICO 2015/2016 TITOLO TESI TARGET SCORING NEL LOAN PLACEMENT

Candidato: Dott. Ing. Maurizio Clemente

Relatore: Prof.ssa Veronica Baldisserri

ABSTRACT

Problematica

Un comparatore *web* di tariffe stava riscontrando forti problematiche di sostenibilità economica su un'attività *outbound* di *loan placement*.

La campagna di *telemarketing* consisteva nel contatto di Clienti che avevano richiesto una quotazione per una polizza RC auto sul sito del comparatore al fine di proporre appuntamenti presso le filiali di zona di società finanziarie *partner* al fine di valutare un prestito personale.

La *Redemption* dell'iniziativa commerciale era ampiamente al di sotto della soglia ritenuta economicamente accettabile e sostenibile dalle società finanziarie che stavano ipotizzando, pertanto, la sospensione immediata della campagna o la rinegoziazione al ribasso delle provvigioni.

Obiettivo

Il *call center* Covisian ho proposto la creazione di modello di *Target Scoring* partendo dallo storico accumulato dal comparatore sulla campagna.

Il comparatore ha fornito a Covisian un campione di 104.137 anagrafiche per la creazione del modello e un lotto di 50.000 anagrafiche da profilare per individuare il *target* ottimale da testare prima di passare alla produzione a regime.

Metodo

Le osservazioni con dati mancanti sono state eliminate (*Listwise Deletion*), data l'incidenza molto contenuta (<0,02%) e la numerosità elevata del campione di analisi.

La variabile "Prezzo preventivo" conteneva numerosi *Miscoded data* che sono stati sostituiti tramite *Regression Imputation* (i.e. creazione di un modello di regressione e sostituzione dei dati errati con le stime del modello).

Sono state create delle variabili derivate (e.g. *recency* del contatto) ritenute dagli esperti di dominio notoriamente influenti nelle iniziative di contatto diretto su liste del settore digital.

Le variabili categoriche sono state trasformate in variabili *dummy* (talora, ove necessario, previa accorpamento delle categorie) per preparare i dati alla successiva applicazione delle tecniche statistiche inferenziali.

Il campione è stato quindi suddiviso in due parti: *Training Set* (70%) e *Test Set* (30%).

Il modello di *scoring* è stato creato sul *Training Set* utilizzando la regressione logistica (*link logit*).

Per la selezione delle variabili da includere nel modello è stato utilizzato un approccio *Stepwise Regression*.

La valutazione globale del modello è stata effettuata tramite *Likelihood-ratio test*.

La significatività dei coefficienti è stata verificata mediante il *Wald test* mentre l'adattabilità ai dati e la calibratura del modello sono state valutate con il test di Hosmer-Lemeshow e un *calibration plot*.

La capacità del modello di classificare correttamente è stata appurata tramite curva ROC, *AUC*, *Lift Chart* e *Cumulative Gains Chart*.

Tramite il metodo *Closest-Topleft* è stato individuato il *cutoff* per la selezione del target da contattare ed è stato verificato che rappresentasse la soglia ottimale, cioè quella che consente di massimizzare il margine operativo lordo del comparatore.

Il modello è stato applicato al lotto di 50mila *Lead* da profilare ed è stato quindi implementata una campagna di test.

Conclusioni

La campagna di test ha ottenuto una *Redemption* di 1,953% contro una *Baseline* (modello *random*) di 1,014% permettendo di ottenere un margine operativo lordo del 35% per il comparatore.

Il modello logistico ha consentito di riportare la campagna *outbound* in redditività e sostenibilità economica, sia per il comparatore che per le società finanziarie.

Dal punto di vista di Covisian, la dimostrata efficacia della Data Science applicata allo *scoring* di liste *outbound* rappresenta la vittoria di un modello di *business* che prevede di emergere in un mercato estremamente competitivo non attraverso una cruenta battaglia sui prezzi, impossibile di vincere con l'agguerrita concorrenza *nearshore* e *onshore*, ma fornendo ai propri Clienti servizi e gestioni evolute ad alto valore aggiunto.

Parole chiave

Target Scoring, Regressione Logistica, Logit, Loan Placement, Prestiti Personali, Outbound, Call Center, Business Process Outsourcing, Data Cleaning, Odds ratio, Listwise Deletion, Stepwise Regression, Likelihood-ratio test, Wald test, Hosmer-Lemeshow, Wald test, Confusion Matrix, Lift Chart, Cumulative Gains Chart, ROC Curve, Closest Topleft, AUC.