Master Universitario di I Livello in:
"ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE" – A.A. 2024/2025

Titolo della tesi:
Probabilistic Calibration of Machine Learning Models in Large-Scale Decision Systems

Autore:
Andrea Guarneri

Abstract

Il presente lavoro ha come obiettivo lo studio della calibrazione probabilistica nei modelli di Machine Learning utilizzati nei sistemi decisionali su larga scala, con particolare riferimento a contesti operativi in cui le probabilità prodotte dal modello guidano direttamente decisioni aziendali.

In molti ambiti applicativi, come la gestione della supply chain, il credit risk o la rilevazione delle frodi, i modelli di Machine Learning vengono utilizzati per stimare la probabilità di eventi futuri. Tuttavia, una buona capacità discriminante del modello non garantisce necessariamente che le probabilità predette riflettano correttamente il rischio reale. Questo fenomeno, noto come miscalibration, può portare a decisioni operative subottimali quando le scelte aziendali dipendono da soglie probabilistiche.

Il lavoro analizza questo problema utilizzando il DataCo Supply Chain Dataset, composto da circa 180.000 osservazioni relative a operazioni di supply chain. Dopo una fase di preprocessing e feature engineering, che include trasformazioni delle variabili categoriali tramite one-hot encoding e l'utilizzo di una procedura di 5-fold cross-validation, vengono confrontati tre modelli di machine learning con caratteristiche differenti: Logistic Regression, Random Forest e XGBoost.

La qualità probabilistica dei modelli viene valutata tramite diverse metriche di calibrazione, tra cui Expected Calibration Error (ECE), Maximum Calibration Error (MCE), Brier Score e Log Loss. Successivamente vengono applicate tecniche di post-calibrazione, in particolare Isotonic Regression per i modelli tree-based e Platt Scaling per la Logistic Regression, al fine di correggere eventuali distorsioni nella scala delle probabilità.

I risultati mostrano che, nonostante una capacità discriminante simile tra i modelli (ROC-AUC ≈ 0.73), i modelli ensemble presentano livelli di miscalibration significativamente

più elevati rispetto alla Logistic Regression. L'applicazione delle tecniche di calibrazione riduce l'Expected Calibration Error fino al 92% per Random Forest e all'86% per XGBoost, migliorando significativamente l'affidabilità delle probabilità predette senza alterare la capacità discriminativa del modello.

Infine, il lavoro analizza l'impatto della calibrazione all'interno di un framework decisionale cost-sensitive, in cui le decisioni operative dipendono dai costi associati ai falsi positivi e ai falsi negativi. Attraverso una sensitivity analysis su 192 scenari decisionali, viene valutato il beneficio economico della calibrazione. In uno scenario operativo di riferimento, la calibrazione del modello XGBoost genera un risparmio di circa 18.060 euro sul test set, equivalente a circa 66.700 euro annui su 100.000 ordini, dimostrando come la calibrazione possa generare valore economico concreto.

I risultati evidenziano che la calibrazione rappresenta un passaggio fondamentale per garantire che le probabilità prodotte dai modelli di Machine Learning siano coerenti con il rischio reale, trasformando modelli predittivi in strumenti più affidabili per il supporto alle decisioni aziendali.