
Tre generazioni di Machine Learning ***Un focus sul Business Value***

<https://blog.minitab.com/blog/3-generations-of-machine-learning-models-a-new-focus-on-business-value>

Versione italiana a cura di Luca Biasibetti.

Quando circa un decennio fa si è iniziato a parlare di data science, la creazione di un modello predittivo era un lavoro complicato che solo i programmatori esperti potevano realizzare. Oggi, nuovi strumenti e nuove tecniche stanno spostando l'attenzione dai dettagli di programmazione alla costruzione di modelli che siano robusti e focalizzati sulla creazione di valore aziendale. In questo articolo, forniremo una struttura per spiegare il funzionamento del Machine Learning e mostriamo come stiamo entrando in una terza generazione di funzionalità relative all'apprendimento automatizzato a cui è possibile attingere.

MACHINE LEARNING: L'ARTE E LA SCIENZA DELLA PREDIZIONE

Che cos'è il Machine Learning? Di base, non è altro che una raccolta di algoritmi che consentono di fare previsioni riguardo a qualcosa che è sconosciuto sulla base dei dati noti. Nel loro libro *Prediction Machines*, gli economisti Ajay Agrawal, Joshua Gans e Avi Goldfarb sottolineano che l'intelligenza artificiale non riguarda la creazione di vera e propria intelligenza, ma la capacità delle macchine di fare previsioni, che sono l'input centrale per il processo decisionale.

Come spiegano gli autori, quando si tratta di fare previsioni, le macchine e gli esseri umani hanno punti di forza e di debolezza distinti. Le macchine sono in grado di gestire interazioni complesse, soprattutto in situazioni ricche di dati, mentre gli esseri umani svolgono un lavoro migliore quando è importante comprendere il processo che ha prodotto i dati. I migliori risultati si ottengono quando vengono combinati i punti di forza di macchine ed esseri umani.

L'approccio odierno al machine learning ha le sue radici nella statistica. La regressione lineare, che esiste da più di un secolo, continua a rimanere un'importante forma di apprendimento automatico. Negli ultimi due decenni, le nuove forme di machine learning sono diventate molto pratiche. Possiamo classificare gli approcci moderni al machine learning in tre generazioni.

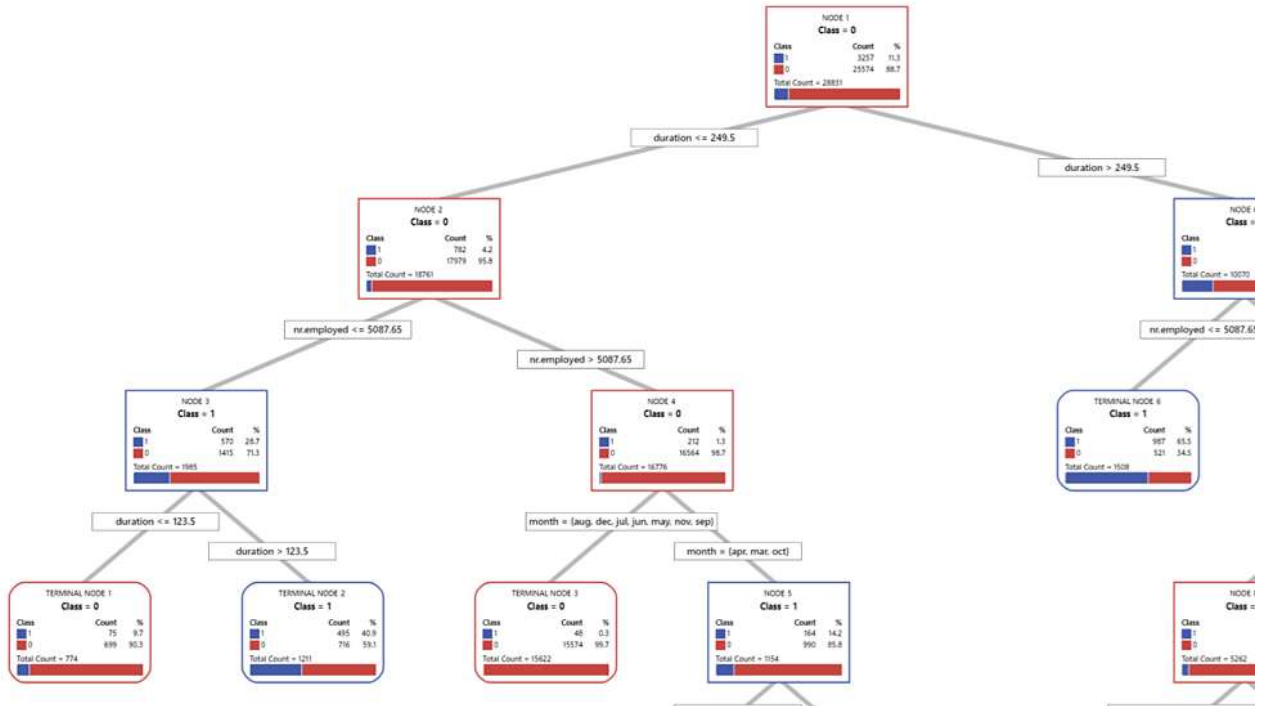
PRIMA GENERAZIONE: REGRESSIONE LOGISTICA, ALBERI DECISIONALI e MINITAB CART

La prima generazione consiste in un insieme di tecniche di modellizzazione di base che sono spesso sufficienti per fare previsioni molto utili. Sono disponibili in diverse tipologie, a cominciare dagli alberi decisionali.

Supponiamo che tu stia cercando di prevedere se un potenziale cliente effettuerà un acquisto in base alla cronologia degli acquisti passata. Entrano in gioco molti fattori altamente intrecciati. Una tipica analisi di Business Intelligence non è in grado di analizzare trame troppo complesse. Un albero decisionale determina sistematicamente, ad ogni passaggio, quale delle variabili disponibili può separare più rapidamente i potenziali clienti che acquistano dai potenziali clienti che non lo fanno.

Il risultato è una tabella di marcia che è possibile seguire per prendere delle decisioni. Nell'esempio mostrato di seguito, generato utilizzando Minitab CART®, la variabile più importante è la durata dell'ultimo contatto con il potenziale cliente. Se il contatto è durato meno

di 249 secondi, c'erano poche possibilità di un acquisto, mentre le probabilità migliorano con contatti più lunghi e così via. Di seguito l'albero decisionale che mostra quanto appena affermato.



[Guarda il nostro Webinar su Minitab e Machine Learning!](#)



[Prova Minitab e CART gratuitamente per 30 giorni!](#)

Altri modelli di seconda generazione includono i Multivariate Adaptive Regression Splines e i Regularized Regression Models, che hanno lo scopo di tenere sotto controllo i problemi relativi alla complessità del modello.

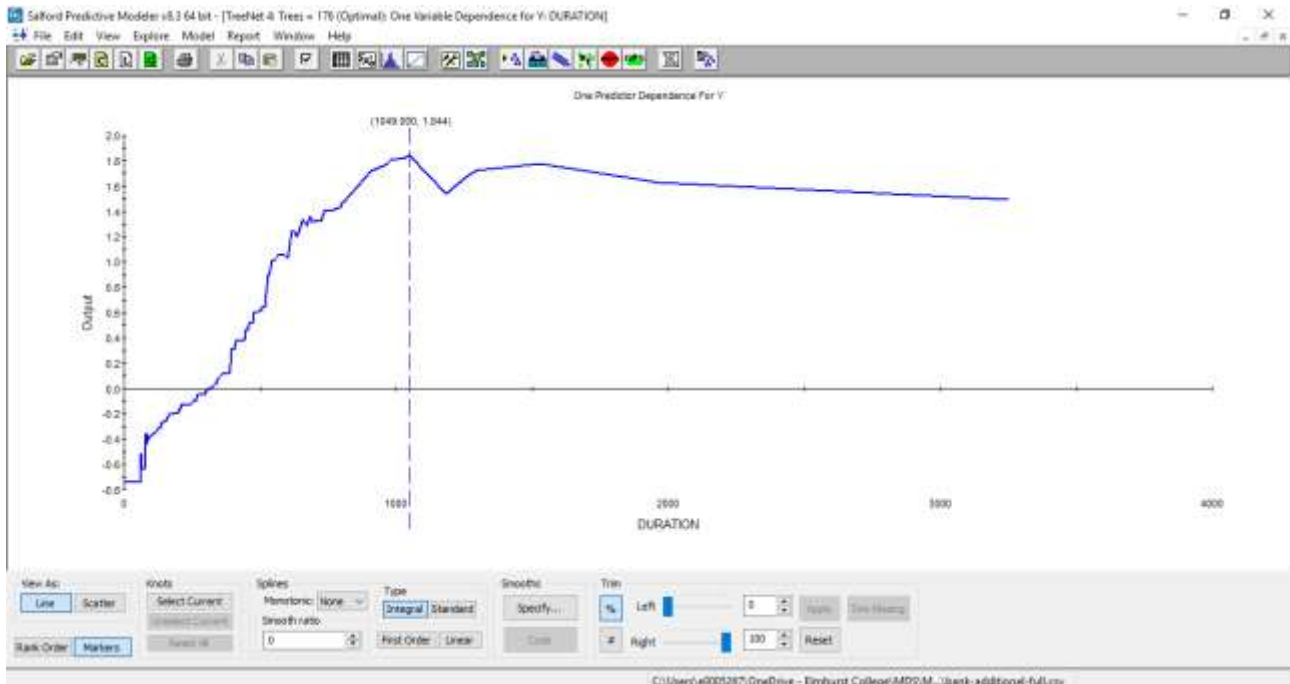
TERZA GENERAZIONE: PRODUZIONE DI UTILI LINEE GUIDA PER IL BUSINESS

Un obiettivo primario, quando si decide di utilizzare modelli di Machine Learning, è quello di produrre informazioni utili per il business. I modelli di Machine Learning di terza generazione estendono la portata delle loro precedenti controparti fornendo nuove informazioni aggiuntive. Per capire meglio, diamo un'ultima occhiata al nostro esempio:

- il nostro modello Random Forest dice che la durata del contatto con il cliente è importante;
- il nostro albero CART dice che la probabilità di acquisto è bassa se il contatto dura meno di 249 secondi.

Non sarebbe interessante sapere qual è la lunghezza ottimale che dovrebbe avere un contatto? A tale scopo può essere utile un algoritmo di terza generazione come *SPM TreeNet® Gradient Boosting*.

Il grafico illustrato di seguito è chiamato diagramma di dipendenza parziale a una variabile e mostra in che modo l'impatto della durata del contatto che incide sulla probabilità di effettuare un acquisto, varia attraverso i valori possibili della durata del contatto stesso. Sembra un'idea semplice, ma in realtà è molto difficile da attuare poiché l'effetto di una variabile come la durata è strettamente correlato agli effetti di molte altre variabili.



Il nostro diagramma di dipendenza parziale ci dice che la prospettiva di concludere positivamente un affare continua a crescere per i contatti che durano fino a circa 1.000 secondi e poi rimangono stabili. Questo è solo un esempio che va oltre l'offerta dei modelli di precedente generazione.

Un'altra area di interesse per i modelli di terza generazione è la capacità di gestire dati di testo e immagini. Questi tipi di dati tendono ad avere, una volta convertiti in numeri, molte più colonne che righe (un incubo per le precedenti tecniche di modellazione). I cosiddetti dati non strutturati come testo e immagini stanno diventando sempre più importanti: a volte portano la maggior parte del potere predittivo di un modello, e quindi la sua capacità di generare valore aziendale. I modelli di terza generazione come GPS e *MARS*[®] hanno le capacità necessarie ad effettuare analisi su questi tipi di dati.

CONCLUSIONE: C'È MOLTO PIU' DI QUESTO DA SAPERE!

I modelli devono essere stabili ed implementabili. Gli strumenti di Machine Learning attualmente utilizzati stanno rendendo possibili entrambi questi risultati. Le sfide, dal punto di vista aziendale, diventano la consapevolezza di ciò che è ora possibile fare e la capacità di utilizzare in modo saggio i potenti risultati dei modelli di machine learning di prima, seconda e terza generazione.

[Hai difficoltà a gestire, comprendere e sfruttare a pieno i tuoi dati? Parla con GMSL](#)

"Portions of information contained in this publication/book are printed with permission of Minitab Inc. All such material remains the exclusive property and copyright of Minitab Inc. All rights reserved."