



CORSO DI FORMAZIONE | ONLINE

## MACHINE LEARNING IN STATA: UN'INTRODUZIONE | MODULO II

Il “*machine learning*” (o “apprendimento automatico”) è un approccio relativamente nuovo all’analisi dei dati, che si colloca nell’intersezione tra statistica, informatica ed intelligenza artificiale. Il suo obiettivo principale è quello di trasformare le informazioni in conoscenza e valore, “lasciando che i dati parlino da soli”.

A tal fine, il *machine learning* limita le ipotesi preliminari sulla struttura dei dati e fa affidamento su una filosofia che supporta lo sviluppo di algoritmi, di procedure computazionali e di ispezione grafica dei risultati più che su assunzioni analitiche e soluzioni algebriche.

Il corso offre, nel [primo modulo](#), una introduzione ad alcune popolari tecniche di *machine learning* utilizzando il software Stata. Stata possiede oggi vari pacchetti per eseguire il *machine learning* che sono tuttavia poco conosciuti da molti suoi utenti. Il programma è stato sviluppato per colmare questa lacuna rendendo i partecipanti più familiari (e meglio informati) sul potenziale di Stata per trarre conoscenza e valore dai dati, possibilmente di grandi dimensioni e “rumorosi”. Nel **secondo modulo**, invece, copre i seguenti temi e metodi: 1) Ricapitolazione delle basi concettuali del *machine learning*; 2) Modelli “locali” di analisi non-parametrica; 3) Modelli “semi-globali” e “globali” di analisi non-parametrica; 4) Alberi decisionali e metodi di analisi “*ensemble*”; 5) Reti neurali.

L’approccio all’insegnamento sarà principalmente basato sul linguaggio grafico e sull’intuizione più che sull’algebra. Le lezioni si avvarranno di esempi sia simulati che reali, e permetterà di bilanciare equamente sessioni teoriche e sessioni pratiche.

Dopo il corso, i partecipanti saranno in grado di utilizzare le tecniche di *machine learning* apprese per varie finalità, in particolare: (i) per studiare relazioni fortemente non-lineari tra *input* ed *output*, identificando la distribuzione degli effetti in un modello libero dalle tradizionali assunzioni della regressione lineare, (ii) sviluppare “regressioni ad albero”, anche a complemento dell’approccio di regressione classico, avvalendosi di utili rappresentazioni grafiche esplicative degli effetti anche in presenza di più fattori predittivi, (iii) sviluppare classificazioni ad alta capacità predittiva utilizzando le reti neurali.

### CODICE CORSO

I-EF35-20L

### DATA E LUOGO

A causa della situazione pandemica ancora in essere, l’edizione 2022 di questo corso di formazione verrà offerta **ONLINE**. Il programma del corso si articola su 2 moduli della durata di 3 ore ciascuno, per complessive 6 ore di lezione oltre 30 minuti di pausa. Le lezioni si svolgeranno nelle mattine del 9-10 Maggio 2022 dalle ore 9.30 alle ore 13.00.

### DESTINATARI

Il corso è di interesse per ricercatori e analisti in economia, medicina, marketing e scienze sociali che desiderano acquisire gli strumenti fondamentali per implementare l’approccio di *machine learning* sui così detti *Big Data*.

### REQUISITI RICHIESTI

Buona conoscenza della statistica ed econometria di base ed in particolare del modello di regressione lineare, delle regressioni *logit/probit* e dell’inferenza classica. E’ consigliata la conoscenza del Software Stata.

Per la partecipazione a questo corso, l’aver seguito il corso precedente ([modulo I](#)) è consigliato ma non strettamente necessario, in quanto nella prima parte verranno ricapitolati i concetti necessari alla comprensione dei metodi ivi trattati.

## PROGRAMMA

### SESSIONE I: LE BASI DEL *MACHINE LEARNING* RICAPITOLAZIONE

1. *Machine Learning*: definizione, logica, utilità
2. La fondamentale non-identificabilità di  $E(y|x)$
3. Misure di bontà di adattamento: *training versus test error*
4. Metodi di ricampionamento e validazione

### SESSIONE II: MODELLI “LOCALI” DI ANALISI NON-PARAMETRICA

1. Oltre i modelli parametrici: una panoramica
2. Approccio locale: logica
3. Metodi locali:
  - Regressione *Kernel*
  - Regressione *nearest-neighbor*
4. Implementazione in Stata

### SESSIONE III: MODELLI “SEMI-GLOBALI E “GLOBALI” DI ANALISI NON- PARAMETRICA

1. Oltre i modelli parametrici: una panoramica
2. Approccio semi-globale e globale: logica
3. Metodi semi-globali
  - *Step function*
  - *Piecewise polynomials*
  - Regressione *spline*
4. Metodi globali
  - Stimatori polinomiali e seriali
  - Modelli parzialmente lineari
  - Modelli additivi generalizzati
5. Implementazione in Stata

### SESSIONE IV: ALBERI DECISIONALI E METODI DI REGRESSIONE “ENSEMBLE”

1. Alberi di regressione e di classificazione
2. Costruire un albero decisionale tramite la “divisione binaria ricorsiva”
3. Potatura ottimale di un albero tramite *cross-validation*
4. Metodi di stima “*ensemble*” basati su alberi decisionali
  - *Bagging, Random Forests e Boosting*
5. Implementazione in Stata

### SESSIONE V: RETI NEURALI

1. Il modello a “rete neurale”: una introduzione
2. Neuroni, strati nascosti e multi-*outcome*
3. Stima di reti neurali
  - Back-propagation tramite l'algoritmo “*gradient descent*”
  - Adattamento con dati *high-dimensional*
4. Validazione dei parametri di una rete neurale
5. Implementazione in Stata



## LETTURE CONSIGLIATE

[Microeconometrics Using Stata](#), Cameron e Trivedi, Revised Edition, StataPress (2010)

The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Hastie, T., Tibshirani, R., Friedman, J., Springer (2009)

An Introduction to Statistical Learning, Gareth, J., Witten, D., Hastie, T., Tibshirani, R., Springer (2013)

"A Super-Learning machine for predicting economic outcomes", MPRA Paper 99111, University Library of Munich, Germany (2020)

## QUOTA DI ISCRIZIONE

La partecipazione al corso è soggetta al pagamento della seguente quota di iscrizione:

Studenti\*: € 355.00

Dottorandi: € 455.00

Università: € 535.00

Commerciale: € 715.00

\*La nostra politica standard è quella di fornire l'accesso alla tariffa "studenti a tempo pieno", a studenti di laurea o master. E' necessario pertanto presentare copia del libretto universitario o un certificato di iscrizione (in carta semplice) all'Università che ne attesti lo status. Studenti di master e dottorandi **part-time**, occupati, dovranno considerare la tariffa riservata alle Università.

I prezzi si intendono IVA 22% esclusa. L'aliquota IVA non sarà applicata per Enti Pubblici soggetti ad esenzione a norma dell'art. 14 c. 10 della L. 537/93 per la partecipazione a corsi di formazione dei propri dipendenti.

La quota di iscrizione include il materiale didattico e una licenza temporanea del software Stata. Dà inoltre diritto ad uno sconto sull'acquisto di una nuova licenza perpetua per singolo utente del Software Stata (ad esclusione della versione per Studenti e Prof+Plan).

L'iscrizione al corso dovrà avvenire tramite lo specifico modulo di registrazione e pervenire a TStat S.r.l. entro il **29 Aprile 2022**. Lo svolgimento è condizionato dal raggiungimento di un numero minimo di 5 partecipanti ed un numero massimo di 8.

Ulteriori informazioni sulla modalità di iscrizione, incluso termini e condizioni di partecipazione sono disponibili nel nostro sito alla pagina <https://www.tstat.it/formazione/intro-machine-learning-stata-20/>.

## CONTATTI

**Monica Gianni**

TStat S.r.l. | Via Rettangolo, 12-14  
I-67039 Sulmona (AQ)  
T. +39 0864 210101

TStat Training | Kleebergstraße, 8  
D-60322 Frankfurt am Main

[formazione@tstat.it](mailto:formazione@tstat.it)

[www.tstat.it](http://www.tstat.it)  
[www.tstattraining.eu](http://www.tstattraining.eu)

