CORSO DI FORMAZIONE | ONLINE

MACHINE LEARNING IN STATA: UN'INTRODUZIONE | MODULO I

CODICE CORSO

I-EF35-1OL

DATA E LUOGO

A causa dell'attuale situazione COVID-19, l'edizione 2021 di questo corso di formazione verrà offerto ONLINE. Il programma è stato suddiviso in 2 sessioni di 3 ore ciascuna nelle giornate del **29-30 Marzo**, dalle 10.00 alle 13.30 con 30 minuti di pausa.

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*.

Gli ultimi anni hanno visto una disponibilità senza precedenti di informazioni su fenomeni sociali, economici e relativi alla salute. Ricercatori, professionisti e responsabili delle politiche hanno oggi accesso a enormi database (i cosiddetti "*Big Data*") su persone, aziende, istituzioni, dispositivi cellulari, web, satelliti, ecc., con sempre maggiore dettaglio.

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 d'ispezione grafica dei risultati più che su assunzioni analitiche e soluzioni algebriche.

Il corso offre 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". Piu specificamente verranno trattati i seguenti temi e metodi: 1) le basi concettuali del *machine learning*, 2) i metodi di ricampionamento e di validazione di un modello, 3) le tecniche di *feature-selection* e specificazione del modello attraverso regressione regolarizzata, 4) le tecniche di *feature-selection* e specificazione del modello attraverso approcci esaustivi e quasi esaustivi, 5) la classificazione con analisi discriminante e con il metodo *nearest-neighbor.*

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 avranno una migliore comprensione del potenziale di Stata per eseguire il *machine learning*, diventando così in grado di padroneggiare compiti di ricerca che includono, tra gli altri: (i) rilevamento d'importanza dei fattori, (ii) estrazione segnale-rumore, (iii) corretta specificazione del modello, (iv) classificazione, sia da un punto di vista di *data mining* che di approccio causale.

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.

https://www.tstat.it/formazione/intro-machine-learning-stata_1ol/

PROGRAMMA

SESSIONE I: LE BASI DEL *MACHINE LEARNING*

1. *Machine Learning*: definizione, logica, utilità
2. Apprendimento supervisionato e non supervisionato
3. Problemi di regressione e di classificazione
4. Inferenza e previsione
5. Errore di campionamento ed errore di specificazione
6. La fondamentale non-identificabilità di E(y|x)
   - Modelli parametrici e non parametrici
   - Il *trade-off* tra accuratezza della previsione e interpretabilità del modello
7. Misure di bontà di adattamento
   - Capacità predittiva "*in-sample*" e "*out-sample*"
   - Il *trade-off* tra distorsione e *variance*
   - La minimizzazione dell'errore quadratico medio
   - *Training-error* vs. *test-error*
   - I criteri di informazione
8. La relazione tra *Machine Learning* ed intelligenza artificiale
9. *Super-learning* e apprendimento dinamico

SESSIONE II: METODI DI RICAMPIONAMENTO E DI VALIDAZIONE

1. Stima del *test-error*
2. Metodi di validazione
   - Approccio con "insieme di validazione"
   - K-fold *cross-validation*
   - Approccio *"leave-one-out"*
3. Metodo *bootstrap*
4. L'algoritmo di *bootstrap*
5. *Bootstrap* vs. *cross-validation* ai fini della valutazione
6. Implementazione in Stata

SESSIONE III: SELEZIONE DEL MODELLO ATTRAVERSO REGRESSIONE REGOLARIZZATA

1. Selezione del modello e corretta specificazione
2. Metodi di regressione *"shrinkage"*
   - Regressione Lasso, Ridge ed elastica
   - Criteri di informazione e *cross-validation* per il Lasso
   - Lasso e inferenza causale
3. Implemetazione in Stata

SESSIONE IV: SELEZIONE DEL MODELLO ATTRAVERSO APPROCCI ESAUSTIVI E QUASI ESAUSTIVI

1. Approccio esaustivo e quasi-esaustivo con criteri di informazione
   - *Best subset selection*
   - *Backward stepwise selection*
   - *Forward stepwise Selection*
2. Implementazione in Stata

SESSIONE V: ANALISI DISCRIMINANTE E CLASSIFICATORE NEAREST-NEIGHBOR

1. Classificatore con analisi discriminante e metodo *nearest-neighbor*
2. Classificatore ottimale Bayesiano e frontiera decisionale
3. Tasso di errore di classificazione
4. Analisi discriminante
5. Analisi discriminante lineare e quadratica
6. Il classificatore *Naive-Bayes*
7. Il classificatore *k-nearest-neighbor*
8. Implementazione in Stata

https://www.tstat.it/formazione/intro-machine-learning-stata-1ol/

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:

Studente*: € 355.00
Università: € 505.00
Commerciale: € 670.00

*Per usufruire dello status "studente" è necessario presentare copia del libretto universitario o un certificato di iscrizione (in carta semplice) all'Università ed essere *studenti a tempo pieno.* Studenti lavoratori 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 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 19 Marzo 2021. 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-1ol/.

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

www.tstat.it
www.tstattraining.eu