

# **Data Mining**

**F8204B014  
F8204B018**

Aldo Solari



# Docente

E-mail : aldo.solari@unimib.it

Pagina web: : <https://aldosolari.github.io/>

Ricevimento : Mercoledì, 17:00-18:00

stanza 2030, edificio U7, II piano



# Prerequisiti

- Probabilità e Statistica Computazionale M
  - Probabilità Applicata
  - Statistica Computazionale
- Statistica Avanzata M
  - Statistica Multivariata
  - Teoria dell'Inferenza Statistica



# Dettagli organizzativi

A partire dalla seconda settimana:

	Lunedì	Martedì	Venerdì
Dove	Lab906	Lab906	Lab906
Inizio	13:00 (3h)	13:00 (3h)	14:30 (2h)



# Obiettivi del corso

- Acquisizione delle tecniche di apprendimento supervisionato (*supervised learning*) e la loro implementazione nell'ambiente di programmazione R
- Forte enfasi su *applied predictive modelling*

## Workload

- $\approx 1/3$  teorico
- $\approx 2/3$  applicato



# Riferimenti bibliografici

- ISL Gareth, Witten, Hastie, Tibshirani (2014) An Introduction to Statistical Learning, with Applications in R. Springer
- AS Azzalini, Scarpa (2012) Data analysis and data mining, an introduction. New York: Oxford University Press
- KJ Kuhn, Johnson (2013) Applied Predictive Modelling. Springer
- WG Wickham, Grolemund (2015) R for Data Science. O'Reilly Cookbooks
- NL Nolan and Lang (2015) Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving. Chapman and Hall/CRC
- ESL Hastie, Tibshirani, Friedman (2009) The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer
- EH Efron, Hastie (2016) Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge University Press



# Materiale didattico

Il materiale didattico è scaricabile da

<https://github.com/aldosolari/DM>

Si precisa che:

- è soggetto a modifiche, quindi si prega di controllare l'ultima versione aggiornata;
- verrà caricato progressivamente, in parallelo allo svolgimento delle lezioni.



# Prova d'esame

- (A) Teoria
  - non è consentito l'uso di libri, appunti o altro materiale
- (B) Programmazione in R
  - è consentito l'uso di materiale didattico fornito dal docente
- (C) Competizioni
  - da svolgere come *homework*



# BeeViva

- Le competizioni si svolgeranno con il supporto di BeeViva s.r.l., uno spin-off dell'Università degli Studi di Padova

<http://www.bee-viva.com/competitions>

- Sempre più spesso il processo di reclutamento di un *data scientist* viene integrato da una prova pratica di analisi dei dati (*business game*)



# Kaggle

Precursore in questo campo è

<https://www.kaggle.com/>

*The Home of Data Science and Machine Learning*



# In Italia

- STATS UNDER THE STARS (8/9/2015)

*Cliente:* Società Italiana di Statistica

*Piattaforma:* Bee-Viva

<http://sus.stat.unipd.it/>

- DATA SCIENTIST RECRUITING GAME (1/10/2015)

*Cliente:* Gruppo PAM

*Piattaforma:* Bee-Viva

<http://www.venetoeconomia.it/2015/09/padova-pam-assume-con-un-gioco-in-ateneo/>

- CALL FOR TALENTS

*Piattaforma:* Mathesia e OPEN SEARCH NETWORK

<http://www.mathesia.com/projects/call-for-talent/>



# In Italia

- BUSINESS GAME FOR DATA SCIENTISTS (19/4/2016)  
*Clienti:* Banco Popolare, ContentWise, Generali, Gruppo AGSM, KPMG, PwC, Salvagnini, Sanmarco Informatica, Unicredit Business Integrated Solutions, Unox  
*Piattaforma:* Bee-Viva  
[http://www.bee-viva.com/competitions/career\\_2016](http://www.bee-viva.com/competitions/career_2016)
- STATS UNDER THE STARS<sup>2</sup>  
*Cliente:* Società Italiana di Statistica  
*Piattaforma:* Bee-Viva  
<http://www.labeconomia.unisa.it/sus2/>
- STATS UNDER THE STARS<sup>3</sup>  
*Cliente:* Società Italiana di Statistica  
*Piattaforma:* Bee-Viva  
<http://local.disia.unifi.it/sus3/>



# Stats under the stars

