

Software Visuali per Analisi Avanzate

Corso sostitutivo di Tirocinio



UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento
DEMM
DIPARTIMENTO DI DIRITTO, ECONOMIA,
MANAGEMENT E METODI QUANTITATIVI

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

1

Calendario del corso Parte 1



UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento
DEMM
DIPARTIMENTO DI DIRITTO, ECONOMIA,
MANAGEMENT E METODI QUANTITATIVI

Giorno	Durata	Argomenti
Venerdì 12/3 16.00 -18.00	2 ore	Presentazione del Corso Il contesto dei Big Data Analitica Avanzata Visual Analytics
Giovedì 18/3 14.00 – 18.00	4 ore	Visual Analytics, Il Quadrante Gartner per la Visual Analytics
Venerdì 19/3 16.00 - 18.00	2 ore	Tableau
Giovedì 8/4 14.00 - 18.00	4 ore	Tableau Design visuale - Ospite Tableau – Ospite
Venerdì 9/4 16.00 - 18.00	2 ore	Tableau Esercitazione finale
Giovedì 15/4 14.00 - 18.00	4 ore	Il Quadrante di Gartner per l'Advanced Analytics Ripasso concetti di Machine Learning Sas: Introduzione, ambiente operativo
Giovedì 22/4 14.00 - 18.00	4 ore	SaS – Lezione a cura di SAS Italia
Venerdì 23/4 16.00 - 18.00	2 ore	SaS – Elaborato Finale

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

2

Comunicazioni di servizio

- Link Slide del corso
 - <https://github.com/valerio75/Software-Visuali-per-Analisi-Avanzate-2020-2021>
- Il secondo modulo del corso

Sommario

Data Science e Machine Learning

Gartner Magic Quadrant Data Science e Machine Learning

Machine Learning

Matrice di Confusione e metriche di classificazione

SaaS

Accedere all'ambiente SaaS

Visual Analytics con SaaS

Pipeline di Machine Learning

I Tipi di analisi

- **Analisi descrittive**, strumenti che permettono di descrivere la situazione attuale e passata di fenomeni (es. processi aziendali). Permettono di accedere ai dati in maniera interattiva (inserendo ad esempio filtri o effettuando operazioni di drill-down) e di visualizzare in modo sintetico e grafico indicatori di prestazione
- **Analisi predittive**, strumenti avanzati che effettuano l'analisi dei dati per rispondere a domande relative a cosa potrebbe accadere nel futuro (sono caratterizzati da tecniche matematiche quali regressione, *forecasting*, modelli predittivi, ecc.)
- **Analisi prescrittive**, strumenti avanzati che, insieme all'analisi dei dati, sono capaci di proporre soluzioni operative/strategiche sulla base delle analisi svolte

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

DSML – Data Science & Machine Learning

- DSML – Data Science & Machine Learning
- Le piattaforme DSML offrono un mix di funzioni di base ed avanzate per la costruzione di soluzioni DSML (principalmente modelli per analisi **predittive** e **prescrittive**).

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

DSML – Data Science & Machine Learning

- Una piattaforma DSML è un prodotto principale ed un portafoglio coerente ed integrato di prodotti, componenti, librerie e framework (proprietarie, di partner e open source) a supporto di una pipeline analitica.
- Gli utenti principali sono professionisti delle data science:
 - expert data scientists
 - citizen data scientists (*)
 - data engineers (**)
 - machine learning (ML) engineers/specialists.

[Magic Quadrant for Data Science and Machine Learning Platforms]

(*) <https://www.forbes.com/sites/stevebanker/2018/01/19/the-citizen-data-scientist/>

(**) <https://cognitiveclass.ai/blog/data-scientist-vs-data-engineer>

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

DSML – Data Science & Machine Learning

- Le piattaforme DSML offrono un mix di funzioni di base ed avanzate per la costruzione di soluzioni DSML (principalmente modelli per analisi **predittive** e **prescrittive**).
- Le piattaforme supportano anche l'incorporazione di queste soluzioni in processi di business, infrastrutture, prodotti ed applicazioni.
- Esse devono supportare i task:
 - Data ingestion
 - Data preparation
 - Data exploration
 - Feature engineering
 - Model creation and training
 - Model testing
 - Deployment
 - Monitoring
 - Maintenance
 - Collaboration

Dott. Ing. Valerio Morfino – Sistemi Visuali per Analisi Avanzate a.a. 2020-2021

DSML – Data Science & Machine Learning

- Non tutte le organizzazioni creano i modelli da zero od in modo completamente autonomo. In alcuni casi è necessario ricorrere ad assistenza e servizi professionali.
- E' quindi importante, ed il Magic Quadrant DSML lo prende in esame, la disponibilità di template ed esempi già pronti nelle piattaforme, perché semplificano l'avvio dei progetti.
- Il Magic Quadrant prende anche in esame la disponibilità di servizi ed assistenza
- Questo aspetto è molto importante: al giorno d'oggi avere progetti di analitica avanzata permette di avere **un vantaggio competitivo!**

DSML – Data Science & Machine Learning

- E' molto importante che le piattaforme supportino non solo la costruzione di un modello, ma anche la sua messa in esercizio. Infatti i benefici di un progetto DSML non potranno essere ottenuti senza che essi:
 - vengano **"embeddati"** nei processi di business e negli ambienti decisionali
 - Vengano mantenuti, monitorati e gestiti nel tempo
- Nonostante vi siano stati numerosi avanzamenti tecnologici nell'ambito dell'AI e del ML, una percentuale elevata di progetti non sono resi operativi.
- Uno dei motivi cruciali è la scarsità di tool che facilitano l'operationalizzazione dei modelli.

Gartner Magic Quadrant

- **Magic Quadrant (MQ)** è una serie di ricerche di mercato pubblicate dalla società di consulenza **Gartner** basati su metodi di analisi proprietari dei dati per mostrare le tendenze del mercato.
- Le analisi sono condotte per diversi settori tecnologici specifici e vengono periodicamente aggiornate.
- Gartner valuta i fornitori in base a due criteri: **completeness of vision** (completezza della visione) e **ability to execute** (capacità di esecuzione)
- Il report di Gartner si divide in 4 quadranti

Gartner Magic Quadrant

- **Leaders** – In questo quadrante sono presenti i vendor che hanno punteggio più alto per Completeness of Vision e Ability to Execute. Hanno quote di mercato, credibilità e le capacità di marketing e di vendita necessarie a guidare la tecnologia al successo. Questi vendor dimostrano chiara comprensione delle necessità del mercato, hanno un pensiero innovativo e piani ben definiti.
- **Challengers** – I vendor di questo quadrante sono presenti nel mercato ed hanno una Ability to Execute buona, tanto da costituire una seria minaccia per i venditori nel quadrante Leader. Hanno prodotti validi, posizione di mercato e risorse sufficientemente credibili per sostenere la crescita continua. Hanno buona redditività finanziaria, ma non hanno le dimensioni e l'influenza dei venditori nel quadrante Leader.

Gartner Magic Quadrant

- **Visionaries** – Un fornitore nel quadrante Visionaries offre prodotti innovativi che affrontano bene i problemi degli utenti finali a livello operativo o finanziario, ma non ha ancora dimostrato la capacità di acquisire quote di mercato o redditività sostenibile. I venditori visionari sono spesso aziende private e obiettivi di acquisizione per aziende più grandi e affermate. La probabilità di essere acquisiti spesso riduce i rischi associati all'adozione dei loro sistemi.
- **Niche Players** – I vendors di questo quadrante sono spesso focalizzati su segmenti di mercato o esigenze verticali specifiche. Questo quadrante può anche includere venditori che stanno riadattando loro prodotti esistenti per entrare nel mercato, o venditori più grandi che hanno difficoltà a far affermare la propria visione.

Magic Quadrant for Data Science and Machine Learning Platforms

Expert data scientists and other professionals working in data science roles require capabilities to source data, build models and operationalize machine learning insights. **Significant vendor growth, product development and myriad competing visions reflect a healthy market that is maturing rapidly.**

https://content.dataiku.com/gartner-mq-2021?__hssc=186155446.ed52e706c5f992d2d6d5c8f3b370a30e.1618430840081.1618430840081.1618430840081.1&__hsfp=1027316311&submissionGuid=6d113b82-c197-4809-9dfb-ed2836dc110c

1618430840082&__hsfp=1027316311&submissionGuid=6d113b82-c197-4809-9dfb-ed2836dc110c



Osservazioni

- Quadrante dei Leader più «affollato» rispetto a quello dell'Analytics and Business Intelligence
- Molte delle aziende presenti in questo Magic Quadrant erano presenti anche in quello dell'Analytics and Business Intelligence
- Alcune aziende di particolare rilevanza:
 - SAS
 - IBM
 - DatalKU
 - Databricks
 - Tibco
 - H2O.ai
 - Microsoft
 - Google

Magic Quadrant for Analytics and Business Intelligence Platforms

Augmented capabilities are becoming key differentiators for analytics and BI platforms, at a time when cloud ecosystems are also influencing selection decisions. This Magic Quadrant will help data and analytics leaders evolve their analytics and BI technology portfolios in light of these changes.



SAS



- SAS offers a variety of software products for analytics and data science supporting statistics, ML, text analytics, forecasting, time series analysis, econometrics and optimization.
- SAS Visual Data Mining and Machine Learning (VDMML) was the core product evaluated for this Magic Quadrant. It incorporates multiple products including Visual Analytics and Visual Statistics.
- SAS provides a dedicated resource center for analytics that offers a series of webinars, events, fact sheets, webcasts, white papers, and more.
- The SAS Resource Center acts as both a library and a support community for users looking for feature information, tips and tricks and case studies in analytics.
- SAS is again positioned as a Leader this year. Its DSML products have a high degree of enterprise readiness and consistently deliver high business value to customers.
- SAS's Ability to Execute continues to be impacted by high license costs, which cause existing and prospective customers to explore other options. It recently launched life cycle product bundles called Unified Insights to reduce licensing complexity.

SaS Viya



- SAS Viya è un'architettura unificata con un ambiente analitico centralizzato per consentire la gestione end-to-end del dato, dall'esplorazione al risultato di business
- SAS è Leader nel Magic Quadrant 2021 di Gartner per le piattaforme di Data Science e Machine Learning.
- SAS Visual Data Mining and Machine Learning fornisce un'interfaccia intuitiva per velocizzare la costruzione di modelli e la generazione di codici.
- Ha funzioni di ricerca dati, feature engineering, riduzione della dimensionalità, analisi esplorativa, modellizzazione, training, tuning e implementazione dei modelli nei processi produttivi.

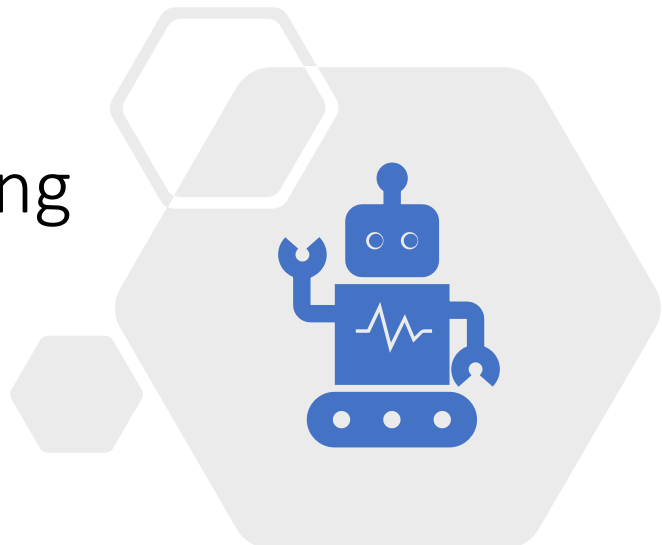
SaS Viya



- Durante il corso utilizzeremo la piattaforma per esplorare le funzionalità di advanced analytics, con particolare riferimento al machine learning supervisionato
- Piattaforma e-learning di SaS per il corso “ Machine Learning Using SAS® Viya”
- Badge pubblicabile su LinkedIn dopo 14 h di lezione e-learning
- Una lezione da SaS Italia
- Possibilità di certificarsi SaS Certified Specialist: Machine Learning Using SAS® Viya

Panoramica sul Machine Learning

Prof. Salvatore
Rampone





21

Hands on!



22

Link utili alla documentazione



- Home del Virtual Learning Environment per SCYP Machine Learning with SAS: <https://vle.sas.com/course/view.php?id=3436>
- Cliccare su: LEARN SAS per accedere all'e-learning (link diretto: <https://vle.sas.com/course/view.php?id=7768>)
- Cliccare su Access SaS Software per accedere a Sas Viya (link diretto: <https://vle.sas.com/course/view.php?id=3436§ion=1>)

La piattaforma SaS - Overview

Machine Learning con SaS

Il problema della Churn Prediction

Accesso alla piattaforma SaS Viya

Creazione di un nuovo progetto

La pipeline di Machine Learning

Visual Analytics con SaS

Primi passi su SaS Viya

- <https://vle.sas.com/mod/scorm/view.php?id=172844>
 - Create a Project
 - Modify the Data Partition
 - Build a pipeline from a Basic Template
- Practice The Demo: Create a Project and load data
- Visual Analytics con Sas Viya
- Nota: Il Dataset da utilizzare è nel folder CPML25

25

Primi passi su SaS Viya

- <https://v4e001.vfe.sas.com/SASDrive/>
- Andare su menù delle applicazioni
- Esplorazione e visualizzazione dei dati
- Cliccare su Dati
- Selezionare COMMSDATA
- Creare grafico a Barre
 - Aggiungere una categoria (osservare cardinalità)
 - Osservare misura di default
- Creare grafico a torta
 - Aggiungere una categoria (osservare cardinalità)
 - Aggiungere una misura
- Dashboard
- Esporta Dati
- Heatmap
- Mappa Ad albero
- Cluster

26

Grazie per l'attenzione



<https://it.linkedin.com/in/valerio-morfino>



vmorfino@unisannio.it