

Master Universitario di I livello in Analisi Dati per la Business Intelligence e Data Science (A.A. 2017/18)

INSEGNAMENTI E DETTAGLIO ARGOMENTI
Software per la Business Intelligence e metodologie di programmazione
5 CFU - 40 ore lezione frontale - 85 ore studio individuale
L'agenda del master e presentazione degli argomenti
Business Intelligence: dal DB Management e Data Quality agli analytics e alla comunicazione dei risultati: la toolbox
Backend: Strumenti per archiviazione e aggiornamento dei dati, flat tables, DW e DataMart, Hadoop e Map reduce per i Big Data
Introduzione agli strumenti: Excel, Qlickview, SAS, Knime, Qgis, Python, SQL
Verifica di fine modulo
Data Management e trattamento dei dati non strutturati
3 CFU - 20 ore lezione frontale - 55 ore studio individuale
Esplorare le fonti informative interne ed esterne all'azienda
"Lessons learned" nella costruzione di basi dati decisionali
Come costruire basi dati decisionali (modellare i dati decisionali, la costruzione della base dati decisionale)
Costruire le basi informative a partire dai bisogni di conoscenza del cliente
Dati pubblici: open data, big data, data governance, metadati, privacy
Introduzione alla Yucca Smart Data Platform (IoT, stream di dati real time) ed esercitazioni di utilizzo
Lavoro di gruppo inerente Data Management, Data Quality e Statistiche introduttive su tematiche inerenti la Pubblica Amministrazione (avvio lavoro in sezione Data Management e conclusione nella verifica modulo Mining I)
Linguaggio SQL
Verifica di fine modulo
Raccolta dei dati e data quality
4 CFU - 32 ore lezione frontale - 68 ore studio individuale
Corso di SAS Base e SAS Guide: Strumenti per la raccolta dei dati e data quality - Access data
Corso di SAS Base e SAS Guide: Strumenti per la raccolta dei dati e data quality - Manage data
Corso di SAS Base e SAS Guide: Strumenti per la raccolta dei dati e data quality - Analyze data
Corso di SAS Base e SAS Guide: Strumenti per la raccolta dei dati e data quality - Reporting
SAS Proc SQL
Introduzione a SAS Data Integration Studio
Esempio di estrazione in laboratorio di informazioni da testi (come articoli di giornale), loro codifica in RDF e interrogazione del triple store risultante tramite SPARQL
Esemplificazione di DATA INTEGRATION
Strumenti per la raccolta dei dati e data quality: Present data
La gestione del data anagrafico per il riconoscimento e la contattabilità del Cliente - non PA
L'assessment di una qualità di una base dati - non PA
Prestare attenzione alla qualità delle informazioni (le dimensioni di qualità dei dati) - PA
Introduzione ai Sistemi Informativi Geografici e uso di Qgis
Verifica di fine modulo

Statistica per la BI
3 CFU - 24 ore lezione frontale - 51 ore studio individuale
Eventi e probabilità
Distribuzioni di probabilità
Campionamento e intervalli di confidenza
Test di ipotesi
Verifica di fine modulo
Mining I – Statistica Descrittiva
3 CFU - 24 ore lezione frontale - 51 ore studio individuale
Uso del software R e dell'IDE Rstudio per analisi statistiche
I dati, le funzioni, i grafici, altri oggetti: statistica descrittiva
Test statistici parametrici e non parametrici
La correlazione e la regressione
ANOVA a una via
Tabelle di contingenza e indipendenza statistica
Verifica di fine modulo
Mining II – Machine Learning e Text Mining
5 CFU - 40 ore lezione frontale - 85 ore studio individuale
Introduzione, preprocessing, classificazione
Esercitazioni su preprocessing e classificazione con Knime
Clustering e pattern
Esercitazioni su clustering con Knime
Sperimentazione con crossvalidation e tuning dei parametri dei modelli
Esercitazioni sul confronto di metodi predittivi con Knime
Natural Language Processing; introduzione (sintassi, semantica); l'interpretazione semantica e il problema della Word Sense Disambiguation
Risorse per l'NLP: WordNet; algoritmi per la disambiguazione
Estrazione di informazioni da testi (come articoli di giornale), estrazione di keywords e riassunto automatico.
rappresentazione della conoscenza e semantic web
semantic Web: strumenti e formati; introduzione al semantic web, con cenni a strumenti e formati.
Verifica di fine modulo
Mining III – Sintesi dell'informazione statistica
3 CFU - 24 ore lezione frontale - 51 ore studio individuale
Individuazione di differenti tipologie di clienti: analisi cluster da un punto di vista operativo
Conoscere i propri clienti costruendone il profilo attraverso la Segmentazione Comportamentale: identificazione dell'obiettivo e impostazione dell'analisi
Sintesi dei fenomeni multidimensionali: analisi fattoriale e analisi delle corrispondenze
Interpretazione dei risultati, passaggio in produzione e supporto alle strategie di marketing dell'azienda
Esempi di sintesi dell'informazione statistica in ambito Pubblica Amministrazione (indicatori e cruscotti di performance, data e text mining, visualizzazione di informazioni)
Verifica di fine modulo

Modelli e modelli di equazioni strutturali
3 CFU - 24 ore lezione frontale - 51 ore studio individuale
Raccogliere e analizzare dati con Python Web Scraping
Regular expressions, cenni
Approfondimenti su Assessment dei Modelli
Modelli di regressione lineare e regressione logistica
Esempi di modelli: inferenziale, confronti di frequenze, speranza di vita, altri eventuali
Introduzione ai modelli non lineari
Modelli strutturali: PATH
Verifica di fine modulo
Tecniche di simulazione
3 CFU - 24 ore lezione frontale - 51 ore studio individuale
Scelta dei modelli (lineari, non lineari)
Stima dei modelli
Tuning dei modelli
Previsioni - What if analysis
Verifica di fine modulo
Interpretazione e comunicazione delle relazioni statistiche
7 CFU - 52 ore lezione frontale - 123 ore studio individuale
Data Visualization
Infografiche
Media, Pubblici e stakeholder della ricerca scientifica: la loro rilevanza nell'impostazione teorica e nella comunicazione dei risultati e nella comunicazione del metodo (probabilismo, falsificazionismo)
Responsabilità legale della ricerca, proprietà dei dati del finanziatore, copy right e open access nella sua diffusione
Informazioni per la Business Intelligence in impresa
I numeri nella comunicazione
Numeracy: valutare i numeri e ciò che ci viene detto con i numeri e con le rappresentazioni grafiche dei numeri
Banco di prova (benchmarking): costruire riferimenti empirici per la valutazione dei numeri
Intentionality: il contesto in cui i numeri vengono prodotti e i modelli "teorici" da cui derivano
Modellizzazione dei processi di comunicazione e valutazione della efficacia
Specificità e problemi della comunicazione dei risultati di ricerche scientifiche e rilevazioni statistiche
Il caso dei Rapporti di sostenibilità economica, sociale e ambientale di imprese e di università
Verifica di fine modulo