



# UNIVERSITÀ DEGLI STUDI DI PALERMO

<b>DIPARTIMENTO</b>	Matematica e Informatica		
<b>ANNO ACCADEMICO OFFERTA</b>	2023/2024		
<b>ANNO ACCADEMICO EROGAZIONE</b>	2023/2024		
<b>CORSO DILAUREA</b>	INTELLIGENZA ARTIFICIALE		
<b>INSEGNAMENTO</b>	FONDAMENTI DI SCIENZA DEI DATI		
<b>TIPO DI ATTIVITA'</b>	B		
<b>AMBITO</b>	50166-Discipline Informatiche		
<b>CODICE INSEGNAMENTO</b>	21616		
<b>SETTORI SCIENTIFICO-DISCIPLINARI</b>	INF/01		
<b>DOCENTE RESPONSABILE</b>	GARLISI DOMENICO	Ricercatore a tempo determinato	Univ. di PALERMO
<b>ALTRI DOCENTI</b>			
<b>CFU</b>	6		
<b>NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE</b>	102		
<b>NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA</b>	48		
<b>PROPEDEUTICITA'</b>			
<b>MUTUAZIONI</b>			
<b>ANNO DI CORSO</b>	1		
<b>PERIODO DELLE LEZIONI</b>	2° semestre		
<b>MODALITA' DI FREQUENZA</b>	Facoltativa		
<b>TIPO DI VALUTAZIONE</b>	Voto in trentesimi		
<b>ORARIO DI RICEVIMENTO DEGLI STUDENTI</b>	<b>GARLISI DOMENICO</b> Mercoledì 15:00 16:30 Stanza 222 presso il Dipartimento di Matematica e Informatica, via Archirafi 34 Venerdì 15:00 16:30 Stanza 222 presso il Dipartimento di Matematica e Informatica, via Archirafi 34		

DOCENTE: Prof. DOMENICO GARLISI

<b>PREREQUISITI</b>	E' necessaria la conoscenza di principi base di analisi, algebra lineare, calcolo delle probabilità, e programmazione.
<b>RISULTATI DI APPRENDIMENTO ATTESI</b>	<p>Conoscenza e capacità di comprensione: Alla fine del corso gli studenti avranno compreso i principali aspetti teorici e pratici della scienza dei dati. Gli studenti conosceranno le varie fasi di un progetto di scienza dei dati e comprenderanno le varie problematiche collegate. Gli studenti avranno preso familiarità con varie tecniche di apprendimento automatico per sfruttare i dati al fine di risolvere compiti complessi. Gli studenti acquisiranno la capacità di applicare metodi e tecniche per la stima, la riduzione della dimensionalità, la classificazione e il clustering di dati di diversa natura.</p> <p>Capacità di applicare conoscenza e comprensione: Gli studenti acquisiranno la capacità di combinare metodi statistici e computazionali per identificare trend nei dati e generare nuova conoscenza.</p> <p>Autonomia di giudizio: Gli studenti saranno in grado di valutare l'adeguatezza di varie soluzioni algoritmiche per un dato contesto applicativo. Essi saranno in grado di valutare l'efficacia dei vari approcci interpretando opportuni indicatori.</p> <p>Abilità comunicative: Gli studenti acquisiranno la capacità di argomentare e comunicare efficacemente le proprie conclusioni usando il linguaggio della scienza dei dati.</p> <p>Capacità d'apprendimento: Gli studenti acquisiranno le competenze necessarie ad apprendere in modo autonomo nuovi metodi e tecniche nell'ambito della scienza dei dati.</p>
<b>VALUTAZIONE DELL'APPRENDIMENTO</b>	<p>Le conoscenze e le competenze acquisite dallo studente saranno verificate attraverso una prova scritta e un colloquio orale. Durante il corso è prevista la realizzare un progetto, o caso di studio, basato sull'analisi di un dataset per cui si richiede un lavoro di gruppo e la produzione di una relazione.</p> <p>Valutazione del progetto La presentazione dei casi di studio verrà valutata secondo i seguenti aspetti: i)Completezza dell'analisi condotta; ii)Trattamento dei dati; iii)Originalità; iv)Capacità di integrazione di codice già noto dalle esercitazioni durante il corso; La valutazione della relazione avrà un peso sul voto finale.</p> <p>Valutazione della prova scritta La prova scritta è costituita da esercizi e domande volti a verificare le conoscenze degli argomenti affrontati durante il corso, e ad applicare le capacità e le conoscenze acquisite. Nello svolgimento assume fondamentale importanza il commento teorico dei risultati ottenuti. L'articolazione della soluzione consente di apprezzare tutti i livelli di preparazione. La valutazione è espressa in trentesimi ed un minimo di 15 è richiesto per accedere alla prova orale.</p> <p>Valutazione per la prova orale La prova orale consiste in un colloquio, volto ad accertare il possesso delle competenze e delle conoscenze disciplinari previste dal corso; la valutazione viene espressa in trentesimi. Durante il colloquio orale lo studente dovrà essere in grado di discutere le soluzioni proposte durante la prova scritta e il progetto; inoltre saranno proposte domande di diverso e crescente livello di complessità al fine di valutare il raggiungimento degli obiettivi formativi e le abilità comunicative dello studente. Infine, allo scopo di valutare l'autonomia di giudizio, sarà richiesto di analizzare le caratteristiche di specifici scenari applicativi e di proporre le soluzioni più adeguate ai problemi individuati. La valutazione finale terrà conto sia del punteggio del progetto, che della prova scritta, così come dell'esito della prova orale.</p>
<b>OBIETTIVI FORMATIVI</b>	Il corso fornisce competenze riguardanti la scienza dei dati. Lo studente imparerà come trarre conclusioni utili da set di dati ampi e diversificati attraverso l'esplorazione, la previsione e l'inferenza. Gli studenti impareranno a porre domande appropriate sui propri dati e a interpretare correttamente le risposte fornite dagli strumenti inferenziali e computazionali. Lo studente prenderà familiarità con tecniche di apprendimento automatico per sfruttare i dati al fine di risolvere compiti complessi. Le esercitazioni comprendono l'implementazione di algoritmi in Python. A tale scopo, una parte del corso sarà dedicata a fornire agli studenti la conoscenza di base di tale linguaggio.
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	Le lezioni sono organizzate utilizzando metodologie di didattica interattiva, attività di apprendimento attivo, e didattica capovolta.
<b>TESTI CONSIGLIATI</b>	Testi principali / main textbooks [dimensionality reduction, linear regression, machine learning with python] :

	<p>Brunton and Kutz, Data Driven Science &amp; Engineering: Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2019. ISBN: 978-1108422093</p> <p>Sebastian Raschka, Machine learning con python. Apogeo. ISBN: 9788850333974</p> <p>Testi per consultazione / reference books [Python, data analysis]:  Blum, Hopcroft, and Kannan, Foundations of Data Science, Cambridge University Press, 2020; ISBN: 978-1108485067  Sinan Ozdemir, Data Science: guida ai principi e alle tecniche base della scienza dei dati, Apogeo. ISBN: 8850334192</p>
--	--

### PROGRAMMA

ORE	Lezioni
4	Python: Programmare con numeri e stringhe. Cicli. Funzioni
2	Esempi di programmazione Python e ambienti di sviluppo.
4	Python: Liste, tuple, insiemi e dizionari. Librerie NumPy e Scikit-learn.
2	Le fasi di un progetto di scienza dei dati: Determinare gli obiettivi. Analisi esplorativa dei dati. Il problema della qualità dei dati: preelaborazione, pulizia dei dati, gestione dei dati mancanti e campionamento.
4	Analisi esplorativa dei dati - Capire i dati: ricerca di relazioni tra dati, verifica di appartenenza a distribuzioni, ricerca di outlier, standardizzazione, covarianza.
2	Costruire buoni set di addestramento. Rimozione e imputazione dei valori mancanti da un dataset. Miglioramento della collocazione in categorie dei dati per gli algoritmi di machine learning. Selezione delle caratteristiche rilevanti per la costruzione di un modello.
4	Riduzione della dimensionalità e regressione lineare. Singular Value Decomposition (SVD). Principal Component Analysis (PCA).
4	Riduzione della dimensionalità: Approssimazione di matrici ed esempi sulla compressione delle immagini.
2	Introduzione al machine learning: Tipi di machine learning. Metodi supervisionati e non supervisionati. Regressione, classificazione, clustering.
4	Apprendimento supervisionato: Classificatore bayesiano, classificatore k-vicini.
4	Apprendimento supervisionato: alberi decisionali
4	Apprendimento non supervisionato: Introduzione al clustering, similarità e dissimilarità, errore quadratico, algoritmi di clustering gerarchici: single link, complete link, average link.
4	Apprendimento non supervisionato: algoritmi di clustering partizionali e k-means.
4	Valutazione dei modelli e ottimizzazione degli iperparametri: Convalida incrociata k-fold. Metodo del "gomito". Metriche per la valutazione delle prestazioni: precision, recall, f-measure. Grafici ROC.