

Denominazione	Data Science
Moduli componenti	-
Settore scientifico-disciplinare	SECS-S/06
Anno di corso e semestre di erogazione	1° anno, 1° semestre
Lingua di insegnamento	-
Carico didattico in crediti formativi universitari	6
Numero di ore di attività didattica assistita	36
Docente	Antonella Nannavecchia
Risultati di apprendimento specifici	<i>Conoscenza e capacità di comprensione</i> Al termine del corso, gli studenti avranno acquisito la conoscenza di tecniche, metodologie e ambienti applicativi per l'acquisizione, l'integrazione, la gestione, l'analisi e la visualizzazione di dataset digitali di grandi dimensioni. Particolare attenzione sarà posta su concetti di <i>predictive analysis</i>, <i>data mining</i>, algoritmi di <i>machine learning</i>, intelligenza artificiale, <i>high performance computing</i>. Saranno utilizzati dataset reali relativi a svariati settori. <i>Capacità di applicare conoscenza e comprensione</i> - Capacità di progettare adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset di grandi dimensioni al fine di trasformare i dati in conoscenza utile a supportare le decisioni strategiche. - Capacità di utilizzare le tecniche e le metodologie acquisite al fine di costruire soluzioni applicative efficienti. - Capacità di implementazione i principali algoritmi di machine learning attraverso l'utilizzo di ambienti e linguaggi di ampia diffusione (linguaggio Python). <i>Autonomia di giudizio</i> Al termine del corso, gli studenti saranno in grado di: - valutare in autonomia adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset di grandi dimensioni, - individuare/scegliere le tecniche e le metodologie più idonee alla soluzione di specifici problemi con implementazione mediante l'utilizzo di ambienti e linguaggi di programmazione. <i>Abilità comunicative</i> Capacità di trasferire e presentare le informazioni in modo chiaro e significativo al fine di supportare efficientemente i processi decisionali.
Programma	Introduzione alla Data Science. Inferenza statistica e modelli di previsione. Introduzione al machine learning. Algoritmi di classificazione. Naive Bayes. Regressione logistica. Linear Discriminant Analysis. K-Nearest Neighbors. Metodi di ricampionamento. Cross-validation. Bootstrap. Alberi decisionali. Alberi di classificazione. Bagging. Random Forest. Boosting. Dimensionality reduction. Unsupervised learning. Principal Component Analysis. K-means. Clustering gerarchico. Support vector machines. Introduzione alle reti neurali artificiali.
Tipologie di attività didattiche previste e relative modalità di svolgimento	L'insegnamento, di 6 CFU, sarà strutturato in lezioni di didattica frontale di 3 ore integrate da esercitazioni in aula informatica. Attraverso le lezioni e le esercitazioni, è favorita la partecipazione attiva degli studenti nell'acquisizione delle tecniche e degli strumenti di Data Science mediante l'utilizzo dei principali ambienti applicativi.
Metodi e criteri di	La valutazione da parte del docente è volta a verificare la conoscenza e la capacità di utilizzo delle

valutazione dell'apprendimento	tecniche e degli ambienti applicativi utilizzati in Data Science. È possibile sostenere l'esame in modalità frequentante o non frequentante. ^[L]_[SEP] Gli studenti frequentanti, che avranno seguito almeno il 75% delle lezioni, potranno sostenere l'esame mediante prova pratica da svolgere su PC. ^[L]_[SEP] La prova consiste nello svolgimento di 4-5 esercizi applicativi e 4-5 domande teoriche. La prova ha una durata compresa tra 1 ora e 1 ora e 30 minuti. Gli esercizi applicativi riguardano la realizzazione di programmi mediante utilizzo del linguaggio Python per l'implementazione dei principali algoritmi di machine learning. Domande teoriche sono volte a verificare la conoscenza acquisita riguardo alle tecniche e agli strumenti della Data Science per la realizzazione di soluzioni applicative efficienti, con particolare attenzione alla progettazione e valutazione di adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset di grandi dimensioni. Gli studenti non frequentanti possono sostenere l'esame in forma orale. Nel corso della prova è accertata la conoscenza acquisita riguardo alle tecniche e agli strumenti della Data Science per la realizzazione di soluzioni applicative efficienti, con particolare attenzione alla progettazione e valutazione di adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset di grandi dimensioni. Durante la prova orale è, inoltre, richiesto lo svolgimento di quesiti applicativi su PC riguardanti la realizzazione di programmi mediante utilizzo del linguaggio Python per l'implementazione dei principali algoritmi di machine learning.
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	La valutazione da parte del docente comporta l'attribuzione di un voto finale espresso in trentesimi. Sul voto finale gli esercizi applicativi pesano orientativamente per 3/5 e le domande teoriche per 2/5. La sommatoria dei voti è posta pari a 31 al fine di attribuire la lode agli studenti che hanno conseguito un voto maggiore di 30.
Propedeuticità	Non sono previste propedeuticità
Materiale didattico utilizzato e materiale didattico consigliato	Testo consigliato: James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to Statistical Learning. New York: Springer. http://faculty.marshall.usc.edu/gareth-james/ISL/ Testi aggiuntivi: Aurélien Géron. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. O'Reilly. https://b-ok.xyz/book/5341845/f49201?regionChanged=&redirect=631890 Jake VanderPlas. (2017) Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly. https://tanthiamhuat.files.wordpress.com/2018/04/pythondatasciencehandbook.pdf