



Denominazione	Data Science
Moduli componenti	-
Settore scientifico-disciplinare	SECS-S/01
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 frontale	36
Docente	Antonella Nannavecchia
Risultati di apprendimento specifici	<p>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>. A tal fine saranno utilizzati <i>dataset</i> reali relativi a svariati settori.</p> <p>Attraverso le nozioni acquisite lo studente acquisirà la capacità di:</p> <ul style="list-style-type: none">● 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;● utilizzare le tecniche e le metodologie acquisite al fine di costruire soluzioni applicative efficienti;● implementazione dei principali algoritmi di machine learning attraverso l'utilizzo di ambienti e linguaggi di ampia diffusione (linguaggio Python). <p>Al termine del corso, gli studenti saranno pertanto in grado di valutare in autonomia adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset di grandi dimensioni, nonché individuare/scegliere le tecniche e le metodologie più idonee alla soluzione di specifici problemi mediante l'utilizzo di ambienti e linguaggi di programmazione.</p> <p>Gli studenti saranno infine in grado di trasferire e presentare le informazioni in modo chiaro e significativo al fine di supportare efficientemente i processi decisionali.</p>
Programma	<p>Saranno trattati i seguenti temi:</p> <ul style="list-style-type: none">● 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.

	<ul style="list-style-type: none"> • Support vector machines. • Introduzione alle reti neurali artificiali.
Tipologie di attività didattiche previste e relative modalità di svolgimento	<p>L'insegnamento sarà strutturato in lezioni di didattica frontale di 3 ore integrate da esercitazioni in aula informatica.</p> <p>Attraverso le lezioni e le esercitazioni è favorita la partecipazione attiva degli studenti nell'acquisizione delle nozioni, delle tecniche e degli strumenti di Data Science mediante l'utilizzo dei principali ambienti applicativi.</p>
Metodi e criteri di valutazione dell'apprendimento	<p>L'esame, che ha complessivamente la finalità di verificare la conoscenza e la capacità di utilizzo delle tecniche e degli ambienti applicativi utilizzati in Data Science, può essere sostenuto in modalità frequentante (per gli studenti che hanno seguito almeno il 75% delle lezioni) o non frequentante.</p> <p>Gli studenti <u>frequentanti</u> potranno sostenere l'esame mediante prova pratica da svolgere su PC. 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.</p> <p>In particolare, gli esercizi applicativi sono volti a verificare la capacità di realizzazione di programmi mediante utilizzo del linguaggio Python per l'implementazione dei principali algoritmi di machine learning.</p> <p>Le domande teoriche sono volte a verificare la conoscenza acquisita e l'abilità comunicativa 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.</p> <p>Nell'ambito della prova viene inoltre valutato l'atteggiamento critico nella scelta delle tecniche e delle metodologie più idonee alla soluzione di specifici problemi.</p> <p>Gli studenti <u>frequentanti</u>, alternativamente alla prova pratica svolta su PC, possono svolgere l'esame in forma orale, con le stesse modalità illustrate per gli studenti non frequentanti.</p> <p>Gli studenti <u>non frequentanti</u> sostengono l'esame in forma orale. Nel corso della prova, che ha una durata indicativa di 60 minuti, 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.</p>
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	<p>La valutazione da parte del docente comporta l'attribuzione di un voto finale espresso in trentesimi.</p> <p>Sia nel caso di prova orale sia nel caso di prova pratica, sul voto finale:</p> <ul style="list-style-type: none"> • gli esercizi applicativi pesano orientativamente per 3/5; si tiene in considerazione nella valutazione sia la capacità di individuare il corretto procedimento, sia la corretta esecuzione del medesimo; • le domande teoriche per 2/5; ai fini della valutazione si considera sia l'evidenza della conoscenza acquisita, sia la proprietà ed articolazione con cui essa è illustrata. <p>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.</p>
Propedeuticità	Non sono previste propedeuticità
Materiale didattico utilizzato e materiale didattico consigliato	<p><u>Testo consigliato:</u> 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/</p> <p><u>Testi aggiuntivi:</u> 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</p>