



<b>Denominazione</b>	Data Science
<b>Moduli componenti</b>	-
<b>Settore scientifico-disciplinare</b>	SECS-S/01
<b>Anno di corso e semestre di erogazione</b>	1° anno, 1° semestre
<b>Lingua di insegnamento</b>	-
<b>Carico didattico in crediti formativi universitari</b>	6
<b>Numero di ore di attività didattica frontale</b>	36
<b>Docente</b>	Riccardo De Blasis
<b>Risultati di apprendimento specifici</b>	<p>Al termine del corso, gli studenti avranno acquisito la conoscenza di tecniche e metodologie per l'acquisizione, l'integrazione, la gestione, l'analisi e la visualizzazione di dataset digitali. Particolare attenzione sarà posta ai concetti di <i>data analysis</i>, <i>data mining</i> e algoritmi di <i>machine learning</i>. A tal fine saranno utilizzati <i>dataset</i> relativi a svariati settori.</p> <p>Attraverso le nozioni acquisite lo studente sarà capace di:</p> <ul style="list-style-type: none"><li>● progettare adeguate soluzioni per l'acquisizione, l'analisi e la gestione di dataset al fine di trasformare i dati in conoscenza utile a supportare le decisioni strategiche;</li><li>● utilizzare le tecniche e le metodologie acquisite al fine di costruire soluzioni applicative efficienti;</li><li>● implementare i principali algoritmi di data analysis e machine learning attraverso l'utilizzo del linguaggio di programmazione Python.</li></ul> <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 dati.</p>
<b>Programma</b>	<ul style="list-style-type: none"><li>● Introduzione alla programmazione con il linguaggio Python</li><li>● Manipolazione e analisi dati con Numpy e Pandas</li><li>● Visualizzazione con Matplotlib</li><li>● Machine learning con SciKit-Learn: saranno trattati i principali algoritmi, tra cui k-nearest neighbor, Naïve Bayes, la regressione lineare e logistica, gli alberi decisionali, le reti neurali e il clustering</li></ul>
<b>Tipologie di attività didattiche previste e relative modalità di svolgimento</b>	L'insegnamento sarà strutturato in 12 lezioni di didattica frontale della durata di 3 ore integrate da esercitazioni in aula informatica.
<b>Metodi e criteri di valutazione dell'apprendimento</b>	<p>L'esame ha la finalità di verificare la conoscenza e la capacità di utilizzo delle tecniche e degli ambienti applicativi utilizzati in Data Science e consiste in una prova da svolgersi al PC.</p> <p>La prova prevede lo svolgimento di esercizi applicativi e domande teoriche aperte ed 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 l'utilizzo del linguaggio Python per l'implementazione dei principali algoritmi di data analysis e machine learning.</p> <p>Le domande teoriche sono volte a verificare la conoscenza acquisita e l'abilità comunicativa riguardo alle tecniche e agli strumenti di Data Science.</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>
<b>Criteri di misurazione dell'apprendimento e di attribuzione del voto finale</b>	<p>La valutazione da parte del docente comporta l'attribuzione di un voto finale espresso in trentesimi.</p> <ul style="list-style-type: none"><li>● 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;</li><li>● 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.</li></ul> <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>
<b>Propedeuticità</b>	Conoscenze necessarie per seguire il corso con profitto: matematica generale, statistica.



<b>Materiale didattico utilizzato e materiale didattico consigliato</b>	Materiale didattico fornito dal docente  Testi consigliati: <ul style="list-style-type: none"><li>• Joel Grus. (2021) Data Science con Python, Dai Fondamenti al Machine Learning. Edizione italiana a cura di Gianluca Moro. EGEA. Edizione inglese Data Science from Scratch. O'Reilly <a href="https://github.com/joelgrus/data-science-from-scratch">https://github.com/joelgrus/data-science-from-scratch</a></li><li>• Peter Bruce, Andrew Bruce, Peter Gedeck. (2020) Practical Statistics for Data Scientists, 2<sup>nd</sup> Edition. O'Reilly <a href="https://github.com/gedeck/practical-statistics-for-data-scientists">https://github.com/gedeck/practical-statistics-for-data-scientists</a></li><li>• Jake VanderPlas. (2016) Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly. <a href="https://jakevdp.github.io/PythonDataScienceHandbook/">https://jakevdp.github.io/PythonDataScienceHandbook/</a></li></ul>
---	--