

|   |   |
|---|---|
| <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>La valutazione dell'apprendimento si basa su un esame orale che consiste nella discussione di un progetto individuale con domande aperte su tutto il programma. Il progetto individuale consiste nello sviluppo di un'analisi di un dataset a scelta dello studente e opportunamente elaborato tramite l'applicazione di modelli e tecniche di Data Science appresi durante il corso. Il progetto dovrà essere sviluppato con il linguaggio di programmazione Python. Maggiori informazioni sul progetto saranno fornite durante il corso.</p> <p>In alternativa al progetto individuale, lo studente può scegliere di sostenere la prova orale con svolgimento di quesiti applicativi su PC in modalità face-to-face e/o domande aperte su tutto il programma. I quesiti applicativi riguardano l'applicazione di modelli e tecniche di Data Science tramite il linguaggio di programmazione Python e la piattaforma Google Colab.</p>  |
| <b>Criteri di misurazione dell'apprendimento e di attribuzione del voto finale</b>  | <p>Il voto finale deriva dalla valutazione complessiva della prova e risulta dalla verifica del livello di raggiungimento dei risultati di apprendimento stabiliti.</p> <p>Il voto finale è espresso in trentesimi e rifletterà la dimostrata conoscenza della materia.</p>   |
| <b>Propedeuticità</b>   | Conoscenze necessarie per seguire il corso con profitto: matematica generale, statistica.   |
| <b>Materiale didattico utilizzato e materiale</b>                                   | Materiale didattico fornito dal docente   |



|                              |  |
|------------------------------|--|
| <b>didattico consigliato</b> | <p>Testi consigliati:</p> <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<br/><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<br/><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.<br/><a href="https://jakevdp.github.io/PythonDataScienceHandbook/">https://jakevdp.github.io/PythonDataScienceHandbook/</a></li></ul> |
|------------------------------|--|