

PERCORSO PROFESSIONALIZZANTE **DIVENTA DATA SCIENTIST**

Le conoscenze avanzate sulle tecniche e sugli strumenti per la gestione, l'analisi, la visualizzazione di dati e per il loro uso con tecniche di machine learning.

4 moduli, fruibili anche singolarmente, solo chi frequenta l'intero percorso può sostenere il test finale.

Prerequisiti: è richiesta esperienza di programmazione in un qualsiasi linguaggio, in particolare di Python; per chi non ha conoscenze è richiesta la partecipazione al Percorso "Programmazione Python".

1° MODULO • GESTIONE E VISUALIZZAZIONE DEI DATI **16, 17 e 18 giugno 2021**

► **Concetti avanzati per la gestione dei dati in Python**

► **Data Modeling and Understanding: libreria Numpy**

- Tipi di dato Numpy
- Operazioni base su vettori e matrici
- Funzioni universali e broadcasting
- Analisi statistiche avanzate
- I/O con Numpy
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Data Management, Processing and Cleaning: libreria Pandas**

- Serie e Dataframe in Pandas
- Indicizzazione e selezione in Pandas
- Operazioni di base in Pandas
- Metodi per la pulizia dei dati
- Combinazione e aggregazione dei dati
- Integrazione con Microsoft Excel
- Integrazione e omogeneizzazione di diversi data set
- Introduzione alle Espressioni Regolari e pulizia del testo in Pandas
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Data Visualization: librerie Matplotlib, Seaborn, Plotly**

- Istogrammi, fitting, bar graph e tendenze
- Serie temporali, intervalli di confidenza e boxplot
- Scatter plot, heatmap e densità di probabilità
- Aggregazione, selezione e sampling
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

▶ 2° MODULO • TECNICHE DI MACHINE LEARNING PER L'ANALISI DEI DATI

23, 24 e 25 giugno 2021

▶ Introduzione al Machine Learning

- Dati di Training, Validation e Test: definizione e creazione
- Metriche di valutazione dell'efficacia dei modelli di Machine Learning
- Ottimizzazione e selezione del modello prodotto
- Overfitting e Underfitting
- Machine learning con la libreria scikit-learn

▶ Metodi non Supervisionati: Clustering

- Introduzione al problema del Clustering
- Tipologie di Clustering: densità, k-means, gerarchico, ecc.
- Similarità tra oggetti ed ottimizzazione
- Clustering con la libreria scikit-learn
 - K-means
 - Density-based Clustering
 - Hierarchical Clustering
- Hands-on ed esercizi (dati numerici e testo) (Aule virtuali individuali e/o a gruppi)

▶ Metodi Supervisionati: Classificazione

- Introduzione al problema di classificazione
- Classificazione binaria e multi-etichetta
- Matrici di confusione e metriche di bontà della classificazione
- Classificazione con la libreria scikit-learn
 - Logistic Regression, kNN
 - Alberi di Decisione
 - SVM
- Hands-on ed esercizi (dati numerici e testo) (Aule virtuali individuali e/o a gruppi)

▶ Metodi Supervisionati: Regressione

- Introduzione al problema di regressione
- Matrici di confusione e metriche di bontà della classificazione
- Regressione con la libreria scikit-learn
 - Regressione Lineare
 - Ridge, Lasso
 - Alberi di Regressione
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)



3° MODULO • TECNICHE DI DEEP LEARNING PER L'ANALISI DEI DATI 12, 13 e 14 luglio 2021

► **Introduzione al Deep Learning**

- Introduzione alle Reti neurali e al Deep Learning
- Reti neurali multi-livello
- Uso della libreria Keras per l'implementazione di Reti Neurali
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Deep Learning per Classificazione**

- Classificazione Binaria e Multi-Classe
- Definizione della loss di una rete per task di classificazione
- Funzioni di attivazione per task di classificazione
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Deep Learning per Regressione**

- Definizione della loss di una rete per task di regressione
- Funzioni di attivazione per task di regressione
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Deep Learning per Dati Complessi**

- Reti convoluzionali e ricorrenti
- Deep Learning per analisi di serie temporali
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

► **Valutazione e Best Practices**

- Valutazione e tuning dei parametri tramite uso del Validation Set
- Overfitting e Underfitting
- Data Augmentation e Dropout per la riduzione dell'overfitting
- Transfer-Learning e generalizzazione
- Hands-on ed esercizi (Aule virtuali individuali e/o a gruppi)

