

PERCORSO DIVENTA DATA SCIENTIST

Percorso pensato per fornire conoscenze avanzate su tecniche e strumenti per la gestione, l'analisi, la visualizzazione di dati e per il successivo loro uso con tecniche di machine learning su tre tematiche principali: clustering, classificazione e regressione.

Per la partecipazione al percorso è richiesta esperienza di programmazione in un qualsiasi linguaggio; per chi non ha conoscenze è richiesta la partecipazione al Percorso Python.



1° MODULO • GESTIONE E VISUALIZZAZIONE DEI DATI

7, 8 e 9 ottobre 2020

► 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

► Data Management, Processing and Cleaning: libreria Pandas

- Oggetti 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
- Hands-on ed esercizi

► 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

PERCORSO DIVENTA DATA SCIENTIST

▶ 2° MODULO • TECNICHE DI MACHINE LEARNING PER L'ANALISI DEI DATI 28, 29 e 30 ottobre 2020

▶ 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 e ottimizzazione
- Clustering con la libreria scikit-learn
 - K-means
 - Density-based clustering
 - Hierarchical clustering
- Hands-on ed esercizi (dati numerici e testo)

▶ Metodi Supervisionati: Classificazione

- Introduzione al problema di classificazione
- Classificazione binaria e multi-etichetta
- Matrice 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)

▶ Metodi Supervisionati: Regressione

- Introduzione al problema di regressione
- Matrice 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

PERCORSO DIVENTA DATA SCIENTIST



3° MODULO • TECNICHE DI DEEP LEARNING PER L'ANALISI DEI DATI **10, 11 e 12 novembre 2020**

► **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

► **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

► **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

► **Deep Learning per Dati Complessi**

- Reti convoluzionali e ricorrenti
- Deep Learning per analisi di serie temporali
- Hands-on ed esercizi

► **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