



Sede Legale
Via Nuova Poggioreale, 61 Torre 7 Piano 9°
C. P.N.INAIL - 80143 Napoli
Tel . 081 7871810 Fax 081 32231125
www.adim.info segreteria@adim.info

TITOLO CORSO: **Data Scientist**

MODALITA' DI EROGAZIONE: <https://www.adimfadminerva.it/>

DURATA: 160 ore

COSTO: 800€

PERIODO / DURATA

La proposta formativa con presenza in aula è articolata in 4 macro aree, sviluppate in 160 ore.

ATTESTATO DI PARTECIPAZIONE

Il percorso di formazione si conclude con una verifica e un questionario di apprendimento del corso di cui è richiesto il completamento per il rilascio dell'attestato di partecipazione.

METODOLOGIE

Il percorso integra la didattica d'aula con una costante interazione con i docenti esperti, per rendere l'apprendimento più concreto ed efficace e più vicino alla realtà operativa vissuta dai partecipanti. Viene proposto l'inquadramento teorico-concettuale dei temi trattati, alternato con esercitazioni, analisi di esperienze, simulazioni pratiche e lavori su best practice.

OBIETTIVO

L'obiettivo della proposta formativa è la creazione di figure professionali di data scientist in grado di inserirsi nell'attuale contesto lavorativo nazionale e internazionale del settore ICT dell'analisi dei dati. La figura professionale del data scientist è molto ambita, poiché ha deve essere il giusto connubio di una serie di competenze multidisciplinari legate fra loro, che vanno dalla statistica, matematica, linguaggi di programmazione, organizzazione dei dati e analisi dei dati.





Sede Legale
Via Nuova Poggioreale, 61 Torre 7 Piano 9°
C. P.N.INAIL - 80143 Napoli
Tel . 081 7871810 Fax 081 32231125
www.adim.info segreteria@adim.info

ABSTRACT

Questo corso si articola su queste competenze imprescindibili che ogni data Scientist deve avere

- Development.
- Data Science.
- Visualization.
- Big Data “framework and platforms” (aspetti legali dei Big Data).
- Machine learning.

Il Data Scientist deve possedere la principale attitudine di ogni scienziato, ovvero la curiosità intellettuale, una fame di conoscenza alimentata dal dubbio virtuoso. In particolare, deve essere interessato a osservare la realtà attraverso una rappresentazione particolare, quella dei dati digitali, senza perdere di vista l’insieme e le relazioni tra mondo virtuale e mondo fisico.

È necessario, inoltre, che il Data Scientist possieda o sviluppi un senso sottile, intuitivo, sinottico relativamente al dominio in cui opera. Dovendo identificare comportamenti emergenti, pattern non banali e dinamiche nascoste, occorre, infatti, che lo scienziato non si limiti all’applicazione delle tecniche e delle conoscenze scientifiche, ma riesca a far proprie le regole non scritte, “intangibili”, dei fenomeni osservati. Questa, d’altra parte, è la caratteristica “immaginifica” che determina la differenza dell’essere umano rispetto alla macchina e che permette di afferrare in modo apparentemente inspiegabile le regole di funzionamento di fenomeni che, se osservati unicamente attraverso la descrizione meccanica, non sembrano essere penetrabili da un avanzamento di conoscenza.

Infine, è necessaria una capacità comunicativa in termini non gergali, che permetta allo scienziato di collegare il linguaggio tipico della scienza con quello dei settori in cui si trova a operare. Il decisore politico, il manager o un pubblico misto devono poter afferrare quanto è stato scoperto attraverso l’analisi dei dati, in modo che la conoscenza estratta non solo diventi patrimonio comune, ma abiliti decisioni basate sull’evidenza e consenta di tradurre il nuovo sapere in azioni capaci di generare impatto misurabile.

Quindi il Data Scientist è il luogo di sintesi di conoscenze tecniche “hard” e di caratteristiche più “umane”, intangibili, che consentono di tradurre in valore aggiunto ciò che altrimenti rimarrebbe una sterile, seppur elegante, indagine.

Per questo la componente di “data humanities” risulta fondamentale nel definire il ruolo e l’importanza del Data Scientist, e segna, a oggi, la principale differenza rispetto all’intelligenza artificiale, non nel descrivere e calcolare, ma nel decodificare l’intrinseca complessità dei fenomeni naturali e sociali, soprattutto quando sono descritti attraverso dati digitali.





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

ATTIVITÀ / FASI DI LAVORO

Macro area 1: Linguaggi di programmazione

Fondamenti di Java

Modulo 1: Introduzione a Java

- Cos'è Java.
- Caratteristiche del linguaggio Java.
- Ambiente di sviluppo.
- Struttura del JDK.
- JShell.
- Integrated Development Environment (IDE).

Modulo 2: Componenti fondamentali di un programma Java

- Cosa significa creare un programma Java.
- Le basi della programmazione object oriented: classi ed oggetti.
- I metodi in Java.
- Le variabili in Java.
- I metodi costruttori.
- Introduzione ai package.
- Altri componenti fondamentali della programmazione Java.

Modulo 3: Stile di codifica, tipi di dati ed array

- Stile di codifica.
- Gestione della memoria.
- Tipi di dati primitivi.
- Tipi di dati non primitivi.
- Introduzione alla libreria standard.
- Gli array in Java.

Modulo 4: Operatori e Gestione del flusso di esecuzione

- Operatori di base.
- Costrutti di programmazione semplici.
- Costrutti di programmazione avanzati.

Modulo 5: Object Orientation





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Incapsulamento e visibilità.
- Ereditarietà ed interface.
- Polimorfismo.
- Eccezioni ed asserzioni.
- I generic ed il collection framework.

Fondamenti di Python

Modulo 1 – Cos'è Python

- Introduzione alla programmazione informatica
- Algoritmi, programmi e linguaggi
- Introduzione a Python
- L'ecosistema Python
- Un po' di storia
- Chi usa Python e per che cosa
- La comunità di Python

Modulo 2 – Primi passi con Python

- Installazione di Python
- IDLE e la shell per interagire con l'interprete
- La filosofia di Python

Modulo 3 – Una calcolatrice programmabile

- Il primo programma: "Hello, World!"
- I commenti in Python
- Personalizzare IDLE
- Valutazione di un'espressione
- Operazioni aritmetiche ed espressioni
- Addizione, sottrazione, moltiplicazione e divisione
- Espressioni aritmetiche e ordine di precedenza
- Potenza, divisione intera e modulo
- Variabili e operatore di assegnamento
- Operatori di assegnamento composto
- Le funzioni predefinite





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Call tip: suggerimenti per le funzioni
- I tipi di dato
- I tipi di dato numerici int e float
- La funzione print()
- Calcolo dell'interesse composto
- Calcolare la velocità
- Stile di programmazione
- Nomi validi e autodocumentanti per gli identificatori
- Il documento PEP8
- Help

Modulo 4 – Sequenza, selezione, operatori relazionali e logici

- La funzione input()
- Somma di due numeri interi inseriti dall'utente
- La sequenza
- Flow chart
- Media aritmetica di due numeri
- Un saluto se...
- Pari o dispari
- La selezione
- Blocchi di codice e indentazione
- Sintassi e comportamento dell'istruzione if...elif...else
- Misuriamo la nostra temperatura
- Operatori relazionali o di confronto
- Il massimo di tre numeri
- Operatori logici e il tipo di dato bool
- Appartenenza a un intervallo
- Lati di un triangolo

Modulo 5 – Istruzioni iterativi e flussi di controllo

- Il ciclo for
- La funzione range()



Certificato
UNI EN ISO 9001:2015
n. CN/12718



Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Il ciclo while
- Cicli annidati .
- Costrutto if.

Modulo 6 – Moduli, funzioni e numeri (quasi) casuali

- La Libreria Standard di Python
- Un pizzico di Pitagora
- Definizione di funzioni
- La stringa di documentazione docstring
- Procedure e funzioni
- Giocando ancora coi numeri
- Numeri primi
- Numeri primi gemelli
- Assegnazione multipla e swap di variabili
- Fibonacci e i conigli
- Spazio dei nomi, import e dir() `__name__` e `'__main__'`
- Variabili locali e globali
- Numeri pseudo-casuali
- Indovina il numero
- Costanti e flag
- Policoriandoli
- Animazioni
- Flying birds

Modulo 7 – Stringhe, liste, tuple, insiemi e dizionari

- Tipi di dato semplici e composti
- Il tipo di dato str
- La codifica dei caratteri: ASCII e Unicode
- ASCII Art e parametri delle funzioni
- Affettare le stringhe
- Formattare le stringhe
- Frase inversa





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Palindromia
- La crittografia
- L'alfabeto farfallino
- Cesare e i messaggi nascosti
- Le liste
- Il tipo di dato list
- Operazioni e funzioni sulle liste
- Liste multidimensionali
- Matrici, quadrati magici e quadrati perfetti
- Il tipo di dato tuple
- Il tipo di dato set
- Il tipo di dato dict
- Un sondaggio inaspettato

Modulo 8 – Errori, test e debugging

- Errori ed eccezioni
- Area e perimetro di un cerchio
- Tipi di errore
- Il costrutto try...except
- L'anno del sorpasso
- Bug e debugging
- Un print per scoprire, un print per capire
- Il debugger integrato di Python
- TDD, Test Driven Development
- Calcolo del fattoriale

Modulo 9 – File, dati e statistica

- Elementi di statistica
- La media
- Frequenze assolute e relative
- Diagrammi
- La moda



Certificato
UNI EN ISO 9001:2015
n. CN/12718



Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- La mediana
- Cenni di calcolo numerico
- Rappresentazione dei numeri nel computer
- Equazione di secondo grado
- Metodo Monte Carlo
- Calcolo del valore di pi greco
- I file
- Gestione di file di testo in Python
- Linguistica computazionale
- Inglese o italiano?
- Caratteri e parole
- Un indice analitico
- La nuvola dei promessi sposi: word cloud

Modulo 10 – Grafica e immagini

- Immagini digitali
- Rappresentazione dei colori
- Digitalizzazione delle immagini
- Immagini raster
- Immagini vettoriali
- Diritto d'autore e licenze software
- Pubblico Dominio
- PIL, pip & Pillow
- Elaborazione immagini con Pillow
- Applicazioni con interfaccia GUI
- Un Ciao, Python! grafico
- Programmazione basata sugli eventi

Macro area 2: Piattaforme di Big Data - storage

Modulo 1: Introduzione a Big Data

- Le tecnologie dei Big Data.
- Acquisizione, Immagazzinamento e organizzazione.





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Integrazione.
- Analisi e tecnologie hardware.

Modulo 2: Introduzione a HADOOP

- Che cos'è HADOOP ?
- I sistemi distribuiti ed HADOOP.
- Confronto fra SQL ed Hadoop.
- Introduzione al MapReduce.

Modulo 3: Iniziare con HADOOP

- Lavorare con i Blocchi di HADOOP.
- NameNode, DataNode, SecondaryNameNode, JobTracker, TaskTracker.
- Eseguire HADOOP.
- Integrazione con gli strumenti RAD.

Modulo 4: I componenti di HADOOP

- Lavorare con I file in HDFS.
- HADOOP data type.
- Mapper, Reducer, Partitioner, Combiner.
- Lettura e scrittura con InputFormat ed OutputFormat.
- Organizzazione dei componenti.

Macro area 3: Piattaforme di Big Data – analisi dei dati

Modulo 1: Scrivere un semplice programma con MapReducer

- Recuperare i data set.
- Costruire un basic template di un programma MapReduce.
- Adattarsi ai cambiamenti delle API di Hadoop.
- Streaming in Hadoop.

Modulo 2: Aspetti avanzati del MapReduce

- Lavorare con job del MapReduce.
- Joining di dati data da differenti sorgenti.
- Creare un Bloom filter.
- Sviluppo di un programma MapReduce complesso.
- Monitoring ed debugging.
- Tuning delle prestazioni.

Modulo 3: Lavorare con HADOOP





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Configurazione dei valori dei parametri di HADOOP.
- Monitoring dello stato del sistema.
- Configurazione dei permessi.
- Assegnare le quote.
- Abilitare il trash.
- Aggiungere DataNode.
- Rimuovere DataNode.
- Manutenzione dei NameNode e dei SecondaryNode.
- Recovering da un failed NameNode.

Modulo 4: Lavorare con AWS

- Introduzione ai servizi in cloud di aws.
- S3 per lo storage.

Modulo5: Lavorare con GCP

- Introduzione ai servizi in cloud di aws.

Macro area 4 – Machine Learning

Modulo 1: Dare ai computer la capacità di apprendere i dati

- Costruire macchine intelligenti per trasformare i dati
- in conoscenza
- I tre diversi tipi di machine learning
- Effettuare previsioni sul futuro
- grazie all'apprendimento con supervisione
- Risolvere problemi interattivi con l'apprendimento
- di rafforzamento
- Scoprire le strutture nascoste con l'apprendimento
- senza supervisione
- Introduzione alla terminologia e alla notazione di base
- Una roadmap per la realizzazione di sistemi
- di apprendimento automatico
- Pre-elaborazione: dare una “forma” ai dati
- Addestramento e selezione di un modello predittivo
- Valutazione dei modelli e previsione su istanze di dati
- Il soft computing





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Logica fuzzy
- Cosa è un sistema esperto ?
- Sistemi esperti con logica fuzzy
- Il paradigma di sviluppo di SE fuzzy
- Peculiarità dei SE neuro fuzzy in ambiente aziendale
- Usare Python per attività di machine learning

Modulo 2: Addestrare gli algoritmi a compiti di classificazione

- Neuroni artificiali
- Implementazione in Python di un algoritmo di apprendimento
- Addestrare un modello del perceptron sul dataset Iris
- Neuroni adattativi lineari e convergenza dell'apprendimento
- Minimizzare le funzioni di costo con la discesa del gradiente
- Implementazione di un neurone lineare adattativo in Python
- Machine learning su larga scala e discesa stocastica
- del gradiente

Modulo 3: AI classificatori di machine learning di scikit-learn

- Scelta di un algoritmo di classificazione
- Primi passi con scikit-learn
- Addestramento di un perceptron tramite scikit-learn
- Modellazione delle probabilità delle classi tramite
- la regressione logistica
- Concetti intuitivi e probabilità condizionali
- della regressione logistica
- I pesi della funzione di costo logistico
- Addestramento di un modello a regressione logistica
- con scikit-learn
- Risolvere l'overfitting tramite la regolarizzazione
- Classificazione a massimo margine con le macchine
- a vettori di supporto
- Individuazione del massimo margine
- Il caso separabile non linearmente utilizzando variabili slack
- Implementazioni alternative in scikit-learn
- Soluzione di problemi non lineari utilizzando una SVM kernel





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Uso della tecnica del kernel per trovare iperpiani
- di separazione in uno spazio di maggiori dimensioni
- Apprendimento ad albero decisionale
- Massimizzare il guadagno informativo: la massima sostanza
- al minimo costo
- Costruire un albero decisionale
- Combinare sistemi di apprendimento deboli e forti
- tramite foreste casuali
- I k vicini più prossimi: un algoritmo di apprendimento pigro

Modulo 4: Costruire buoni set di addestramento: la pre-leaborazione

- Il problema dei dati mancanti
- Eliminazione dei campioni e delle caratteristiche con valori mancanti
- Imputazione dei valori mancanti
- Funzionamento dell'API di stima di scikit-learn
- Gestione di dati categorici
- Mappaggio di caratteristiche ordinali
- Codifica delle etichette delle classi
- Esecuzione di una codifica one-hot
- su caratteristiche nominali
- Partizionamento di un dataset nei set di addestramento e di test
- Portare tutte le caratteristiche sulla stessa scala
- Selezione delle caratteristiche appropriate
- Soluzioni sparse con la regolarizzazione L1
- Algoritmi sequenziali per la selezione delle caratteristiche
- Valutazione dell'importanza delle caratteristiche
- con le foreste casuali

Modulo 5: Compressione dei dati tramite la riduzione della dimensionalità

- Riduzione della dimensionalità senza supervisione tramite l'analisi del componente principale
- Varianza totale e spiegata
- Trasformazione di una caratteristica
- L'analisi del componente principale in scikit-learn
- Compressione dei dati con supervisione, tramite l'analisi
- discriminante lineare (LDA)





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Calcolo delle matrici a dispersione
- Selezione dei discriminanti lineari per il nuovo sottospazio delle caratteristiche
- Proiezione dei campioni sul nuovo spazio di caratteristiche
- Analisi LDA con scikit-learn
- Uso della kernel PCA per il mappaggio non lineare
- Funzioni kernel e tecniche kernel
- Implementazione di una kernel PCA in Python
- Proiezione di nuovi punti di dati
- Kernel PCA con scikit-learn

Modulo 6: Valutazione dei modelli e ottimizzazione degli iperparametri

- Accelerare il flusso di lavoro
- Caricamento del dataset relativo al cancro al seno nel Wisconsin
- Combinare i trasformatori e gli estimatori in una pipeline
- Uso della convalida incrociata k-fold per valutare le prestazioni del modello
- Il modello holdout
- La convalida incrociata K-fold
- Debugging degli algoritmi con le curve
- di apprendimento e di convalida
- Diagnosi dei problemi di bias e varianza con le curve di apprendimento
- Soluzione dei problemi di overfitting e underfitting
- con le curve di convalida
- Ottimizzazione dei modelli di machine learning
- tramite ricerca a griglia
- Ottimizzazione degli iperparametri tramite la ricerca a griglia
- Selezione dell'algoritmo con convalida incrociata nidificata
- Varie metriche di valutazione delle prestazioni
- Lettura di una matrice di confusione
- Ottimizzazione della precisione e del recall di un modello di classificazione
- Tracciamento di un ROC
- Le metriche di valutazione per la classificazione multiclasse
- Sviluppo del SE
- Analisi di fattibilità
- Identificazione dei casi normali





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Identificazione dei casi patologici
- Uso dell'applicazione
- Verifica dei risultati

Modulo 7: Combinare più modelli

- Apprendimento d'insieme
- Implementazione di un semplice classificatore
- con voto a maggioranza
- Combinare algoritmi differenti per la classificazione
- con voto a maggioranza
- Valutazione e ottimizzazione del classificatore d'insieme
- Bagging: costruire un insieme di classificatori
- da campioni di bootstrap
- Sfruttare i sistemi di apprendimento deboli tramite
- un boost adattativo

Modulo 8: Progetto Flow e presentazione eccellente

- Accedere al dataset delle recensioni dei film di IMDb
- Introduzione al modello bag-of-words
- Trasformazione delle parole in vettori di caratteristiche
- Valutazione della rilevanza delle parole in base
- alla frequenza (inversa) dei termini nel documento
- Pulitura dei dati testuali
- Elaborazione dei documenti in token
- Addestramento di un modello a regressione logistica
- per la classificazione dei documenti
- Lavorare su grossi insiemi di dati: algoritmi online
- e apprendimento out-of-core
- Presentazione del caso
- Il modello
- Il questionario

Modulo 9: Embedding di un modello in un'applicazione web

- Serializzazione di uno stimatore scikit-learn non addestrato
- Impostazione di un database SQLite per la memorizzazione dei dati
- Sviluppare un'applicazione web con Flask





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Una prima applicazione web con Flask
- Convalida e rendering del modulo
- Trasformazione del classificatore di film in un'applicazione web
- Pubblicazione dell'applicazione web su un server pubblico
- Aggiornamento del classificatore delle recensioni di film

Modulo 10: Previsioni di variabili target continue: l'analisi a regressione

- Introduzione a un modello a regressione lineare semplice
- Esplorazione del dataset Housing
- Visualizzazione delle caratteristiche importanti di un dataset
- Implementazione di un modello a regressione lineare OLS
- Risoluzione dei parametri di regressione nella discesa del gradiente
- Stima del coefficiente di un modello regressione lineare tramite scikit-learn
- Adattamento di un solido modello a regressione utilizzando RANSAC
- Valutazione delle prestazioni dei modelli a regressione lineare
- Uso di metodi regolarizzati per la regressione
- Trasformare un modello a regressione lineare in uno a regressione a curva polinomiale
- Modellazione di relazioni non lineari nel dataset Housing
- Risoluzione delle relazioni non lineari tramite foreste casuali

Modulo 11: Lavorare con dati senza etichette: l'analisi a cluster

- Raggruppare gli oggetti per similarità utilizzando l'algoritmo k-means
- K-means++
- Clustering hard e soft
- Il metodo Elbow per trovare il numero ottimale di cluster
- Quantificare la qualità del clustering tramite grafici a silhouettes
- Organizzazione dei cluster come un albero gerarchico
- Esecuzione di un clustering gerarchico su una matrice delle distanze
- Collegamento dei dendrogrammi a una mappa termica
- Applicazione del clustering agglomerativo tramite scikit-learn
- Individuazione delle regioni a elevata densità tramite DBSCAN

Modulo 12: Reti neurali artificiali per il riconoscimento delle immagini

- Modellare funzioni complesse con reti neurali artificiali
- Ripasso sulle reti neurali monolivello
- Introduzione all'architettura di una rete neurale multilivello





Sede Legale

Via Nuova Poggioreale, 61 Torre 7 Piano 9°

C. P.N.INAIL - 80143 Napoli

Tel . 081 7871810 Fax 081 32231125

www.adim.info segreteria@adim.info

- Attivazione di una rete neurale tramite propagazione in avanti
- Classificazione di cifre scritte a mano
- Procurarsi il dataset MNIST
- Implementazione di un perceptron multilivello
- Addestramento di una rete neurale artificiale
- Calcolare la funzione logistica di costo
- Addestramento delle reti neurali tramite retropropagazione
- Aspetti intuitivi della retropropagazione
- Debugging delle reti neurali con il controllo dei gradienti
- Convergenza nelle reti neurali
- Altre architetture di reti neurali
- Reti neurali convolutive
- Reti neurali ricorrenti
- Un'ultima parola sull'implementazione delle reti neurali

