

# Python for financial players

## Percorso di Professional

17 ottobre – 4 dicembre 2024

6 febbraio – 13 marzo 2025

# Overview

Python è un linguaggio di programmazione attualmente utilizzato in molteplici ambienti, la sua recente diffusione ne fa uno dei linguaggi fondamentali da apprendere e che ogni sviluppatore dovrebbe conoscere. Python è un linguaggio semplice e immediato ed è ideale per essere usato in molteplici applicazioni, in particolar modo per il Machine Learning oppure per quanto riguarda i nuovi trend tecnologici come l'Internet of Things e l'Intelligenza Artificiale (IA). L'analisi dei dati incontra il linguaggio di programmazione Python grazie alla sua versatilità e alle librerie specializzate. Python è noto per la sua sintassi semplice e leggibile, il che lo rende un linguaggio di programmazione accessibile anche ai principianti, facilitando l'ingresso nel campo dell'analisi dei dati. Python è ampiamente utilizzato per lo sviluppo web, l'analisi dei dati, l'automazione e l'apprendimento automatico. Le numerose librerie di analisi dati di Python come NumPy, Pandas, Matplotlib e Seaborn, semplificano l'importazione, la

manipolazione e la visualizzazione dei dati, accelerando il processo di analisi. Python è anche il linguaggio principale per l'implementazione di algoritmi di Machine Learning e Intelligenza Artificiale. Il Percorso offre tutte le competenze necessarie per programmare in Python, i partecipanti saranno guidati nell'utilizzare tecniche avanzate come l'uso di funzioni specifiche, gli scope delle variabili, ecc. Infine, saranno trattate le basi di programmazione Python e verranno affrontati i fondamenti della programmazione stessa con una introduzione al linguaggio.

CRIF Academy propone un percorso formativo per il mercato finance orientato allo sviluppo del Capitale Umano per l'adozione delle nuove tecnologie e metodologie dell'analisi dati e della modellazione attraverso approcci classici e innovativi, quali Machine Learning e Intelligenza Artificiale applicati ai processi del banking.

## Il Percorso Python for Financial Players 2024 è composto da 2 moduli

### MODULO A

#### Fondamenti di Python

Il modulo A è volto a fornire le basi del linguaggio Python abilitando i partecipanti all'analisi del dato e nel condurre le prime analisi statistiche sul campione.

### MODULO B

#### Machine Learning

Il modulo B si basa sull'utilizzo di Python per lo sviluppo di modelli data-driven, da quelli lineari a quelli Machine Learning, attraverso sessioni di hands-on.



# Modulo A

## Fondamenti di Python

17 ottobre – 4 dicembre 2024

### PRACTICE LAB

Ogni modulo sarà accompagnato da un laboratorio pratico conclusivo, progettato per mettere in atto la parte teorica.

## A.1

### Introduzione a Python e alla Data Science

17 ottobre 2024 | 9:30 – 13:30

**Overview della Data Science nell'ambito del rischio di credito e in particolare per gli sviluppi di modelli dedicati al rischio di credito (modelli PD).**

1. Comprendere i settori che utilizzano l'analisi avanzata dei dati
2. Indagare l'impatto che l'analisi dei dati può avere sul processo decisionale
3. Spiegare come i professionisti dei dati preservano la privacy e l'etica dei dati
4. Sviluppare un piano di progetto considerando ruoli e responsabilità dei membri del team
5. Overview sulle caratteristiche dell'ambiente di laboratorio
6. Nozioni di base sulla sintassi di Python (Variabili, tipi dati, Operatori, Commenti)
7. Strutture dati (liste, tuple, dizionari) e relativi operatori
8. Operatori logici e di controllo (for, if, while)
9. Definizioni di funzioni e funzioni anonime (lambda function)

## A.2

### Gestione di file e moduli

24 ottobre 2024 | 9:30 – 13:30

**Basi di manipolazione, accesso e scrittura di file e principali elementi organizzativi di un codice Python nell'ambito della data science.**

1. Lettura e scrittura di file di testo
2. Operazioni di file (apertura, chiusura, lettura/scrittura)
3. Gestione dei percorsi di file
4. Importazione di moduli
5. Creazione di moduli personalizzati
6. Organizzazione dei moduli in pacchetti

## A.3

---

### Data Preparation

---

7 novembre 2024 | 9:30 – 13:30

**Strumenti per preparare e manipolare i dati, con esempi e focus su dati comuni nello sviluppo di un modello di PD.**

1. Creazione e manipolazione di un database
2. Lettura e scrittura di file (csv, txt, xlsx, sas7bdat, parquet)
3. Selezione, filtraggio e ordinamento di un Dataframe
4. Manipolazione del Dataframe (drop di colonne righe, nomenclatura ecc.)
5. Join e Union di dataframes
6. Gestione di variabili categoriche (creazione dummies, colonne categoriche)

## A.5

---

### Data Visualization (plot, grafici)

---

4 dicembre 2024 | 9:30 – 13:30

**Il modulo mira a fornire i fondamentali per gli strumenti di visualizzazione del dato con esempi coerenti dell'analisi esplorativa del dato nell'ambito del rischio di credito.**

1. Introduzione alle librerie di visualizzazione del dato (matplotlib, Seaborn)
2. Creazione di grafici di base (plot, scatter plot, istogrammi, box-plot)
3. Personalizzazione dei grafici (label, titoli, font, colori)
4. Esportazione dei grafici in immagine (png, jpeg, tiff)

## A.4

---

### Analisi esplorativa dei dati (EDA)

---

20 novembre 2024 | 9:30 – 13:30

**Strumenti per la preparazione del dato e gestione duplicati, outlier di valori mancanti e analisi tipiche della modellazione del credito.**

1. Pulizia e preparazione del dato
2. Gestione dei valori mancanti e duplicati
3. Gestione degli outlier
4. Statistiche descrittive
5. Quantili e Binning

# Modulo B

## Machine Learning

6 febbraio – 13 marzo 2025

### PRACTICE LAB

Ogni modulo sarà accompagnato da un laboratorio pratico conclusivo, progettato per mettere in atto la parte teorica.

#### B.1

---

##### Introduzione al Machine Learning

---

6 febbraio 2025 | 9:00 – 13:00

**Introduzione del Machine Learning e della sua importanza nei modelli di credito e relative sfide. Panoramica sul workflow di sviluppo e dei principali tool di Python.**

1. Tipi di apprendimento (Supervisionato, non-supervisionato)
2. Tipologia di problemi di classificazione
3. Workflow di un progetto di Machine Learning
4. Panoramica delle librerie di machine learning (scikit-learn)
5. Suddivisione del campione di train e test

#### B.2

---

##### Feature engineering & selection

---

13 febbraio 2025 | 9:00 – 13:00

**Analisi del processo di machine learning.**

1. Creazione del campione di train e test (Validazione)
2. Creazione di Feature custom
3. Calcolo del Weight-of-Evidence e Information Value
4. Standardizzazione delle variabili
5. Feature encoding (One-Hot-Encoding, Leave-One-Out, WoE)
6. Metriche di valutazione (Confusion, Matrix, ROC-AUC/Gini Index, F1 score)
7. Feature selection monovariata e multivariata



## B.3

---

### Apprendimento supervisionato

---

20 febbraio 2025 | 9:00 – 13:00

**Strumenti per lo sviluppo di rischio di credito (modelli PD) tramite tecniche tradizionali (logistica) ed innovative (machine learnings) e loro valutazione in termini di performance.**

1. Modelli Lineari – Regressione Lineare, Regressione Logistica
2. Modelli ad albero (Alberi Decisionali, Random Forest)
3. Metriche di valutazione (Confusion Matrix, Precision, Recall, ROC – AUC/Gini Index. F1-score)

## B.5

---

### Model Tuning, Validazione e Testing

---

6 marzo 2025 | 9:00 – 13:00

**Nel modulo saranno presentate tecniche atte a prevenire l'overfitting (in linea con le aspettative EBA), la calibrazione di modelli e il salvataggio e inferenza del modello addestrato.**

1. Tuning degli iperparametri con tecniche di Grid-Search
2. Validazione (test) del modello
3. Calibrazione di modelli di classificazione binaria
4. Salvataggio del modello su file pickle
5. Inferenza di un modello addestrato

## B.4

---

### Apprendimento non supervisionato

---

27 febbraio 2025 | 9:00 – 13:00

**Il modulo fornirà gli strumenti per lo sviluppo di modelli di apprendimento automatico non supervisionato.**

1. Modelli di clustering – (KMEANS, Cluster Gerarchico, DBSCAN)
2. Modelli di misture di Gaussiane
3. Modelli di Anomaly Detection (Isolation Forest, One-Class SVM)
4. Metriche di valutazione del clustering

## B.6

---

### Model Explainability

---

13 marzo 2025 – 9:00 – 13:00

**Il modulo fornirà l'importanza della spiegabilità dei modelli nell'ambito del Machine Learning e in linea con le aspettative regolamentari di EBA e quelle dell'Artificial Intelligence (AI-ACT).**

1. L'importanza dell'interpretabilità dei modelli per prendere decisioni consapevoli
2. La spiegabilità del modello da un punto di vista dell'AI-ACT
3. Interpretabilità vs spiegabilità
4. Spiegabilità globale e locale (SHAP, LIME)
5. Applicazione di un caso di esempio

# Calendario

## Modulo A

---

### Fondamenti di Python

---

17 ottobre 2024 | 4 dicembre 2024

Il Modulo A è volto a fornire le basi del linguaggio Python abilitando i partecipanti all'analisi del dato e nel condurre le prime analisi statistiche sul campione.

**A.1** Introduzione alla Data Science

**A.2** Introduzione a Python

**A.3** Gestione di file e moduli

**A.4** Data Preparation

**A.5** Analisi esplorativa dei dati (EDA)

**A.6** Visualizzazione del dato (plot, grafici)

## Modulo B

---

### Machine Learning

---

6 febbraio 2025 | 13 marzo 2025

Il Modulo B si basa sull'utilizzo di Python per lo sviluppo di modelli data-driven, da quelli lineari a quelli Machine Learning, attraverso sessioni di hands-on.

**B.1** Introduzione al Machine Learning

**B.2** Feature engineering & selection

**B.3** Apprendimento supervisionato

**B.4** Apprendimento non supervisionato

**B.5** Model tuning, valutazione e testing

**B.6** Model explainability

# CRIF Academy



Clicca **qui** per iscriverti

Contatti

051 417 5110 | [crifacademy@crif.com](mailto:crifacademy@crif.com)

