

Data Science



תוכנית הלימודים

Part A Python Development (25 Hs)

בחלק זה נלמד את שפת הפיתוח השלטת כיום בעולמות ה Data, שפת ה Python. בחלק זה נבסה לעומק את כל יסודות השפה, החל מנושאי תחביר פשוטים ועד מודולים ייחודיים לשפה זו.

בוגר מודול זה יהיה מפתח Python לכל דבר וענין, ברמה הנדרשת בתעשייה בכלל ובעולמות ה Data בפרט.

Module 1 : Introduction

- Background
- Why Python
- Static vs. Managed Vs. Dynamic Languages
- Usages
- Development Tools

Module 2 : Python basic data types

- Expressions, Statements, Variables
- Working With Numbers
- Working With Strings
- Working With Booleans
- Working With Arrays
- Input & Output

Module 3 : Python common operators

- If Statement
- If-else statement
- If-elif statement
- Switch statement
- While Loops
- For Loops
- Break & Continue
- Range function

Module 4 : Collections

- Working With Lists
- Working With Dictionaries
- Working With Tuples
- Working With Sets
- Sorting Dictionaries
- Copying Collections
- Copying Collections - shallow and deep copy
- Advanced usages with Map, Filter & Reduce

Module 5 : Functions

- Simple functions
- Default Values
- Named Args
- Map
- Filter
- Reduce
- Anonymous Functions
- Built-In Functions
- Scopes : Global , Local , Nested
- Nested Functions
- Lambda Expressions
- Callback functions

Module 6 : Exceptions Handling

- Bug, Error & Exception
- Exception Syntax
- Exception Levels
- The try-catch-finally keywords
- Custom Exceptions
- Raising an Exception

Module 7 : Object Oriented Programming

- Classes
- Objects
- Magic Functions
- Effective documentation using docstrings
- O.O. Concepts
- Fields
- Abstract Data Types
- Inheritance
- Special Methods
- Virtual Functions

Module 8 : Modules & Packages

- What are Modules?
- Importing Modules
- Testing Modules
- Built in Modules
- User Defined Modules
- Search Hierarchy
- Multiple source files
- Byte code
- Packages & PyPi

Final Module Ex

Part B Data Analysis With Python (25 Hs)

בחלק זה נתמקד במגוון מהספריות הנפוצות ביותר של שפת Python לעיבוד ואנליזה של מידע: החל מאיסוף המידע, דרך ניתוח שלו, אופטימיזציה של מודלים והצגה ויזואלית שלו. נכיר שימושים מתמטיים ומדעיים בספריות NumPy, scipy ו Matplotlib.

גם בחלק זה נתרגל את כלל החומר ונפתור תרגיל סיום מודול.

Module 1 : Working with Files & Data

- Accessing files
- Reading files
- Writing to files
- Text files
- XML & JSON files
- Binary Files
- Yaml Files & ruaml_yaml package
- Working with DB
- Working with the Web

Module 2 : The NumPy Library

- Working with Array
- Broadcasting
- N-Dimensional Arrays
- Reading Text Files

Module 3 : The Matplotlib Library

- Matplotlib objects
- Plotting
- The pyplot state machine

Module 4 : The scipy Library

- Useful usages
- Non-Linear equations optimization
- Non-Linear equations solvers
- Root vs. Fsolve techniques

Module 5 : The Pandas Library (If time permits)

- Series & Index
- DataFrame
- Group By

Final Module Ex

Part C Machine Learning (60 Hs)

במודול זה נבסה לעומק את הנושא החם ביותר בעולמות ה Data Science והבינה המלאכותית – Machine Learning. אותו תחום המאפשר למחשבים לפעול "עצמאית" ללא תכנותם.

במודול זה נבסה את כל הטכניקות, הכלים והאלגוריתמים השונים של תחום חשוב זה

Module 1 : Introduction to Machine Learning (ML)

- The ML role on Artificial Intelligence world
- Machine Learning vs. Deep Learning
- ML Algorithms
- Practical use cases
- Methodologies & Tools
- Scikit-Learn
- Pandas
- TensorFlow & PyTorch

Module 2 : ML Model lifecycle

- Exploratory analysis
- Data cleaning
- Feature engineering
- Algorithm selection
- Model training
- What's next ?

Module 3 : Data Preparation

- Data normalization
- Data cleaning
- Data filtering
- Data scaling
- Feature selection
- Feature extraction

Module 4 : Supervised Learning Algorithms

- K-nearest neighbours
- Linear models
- Linear Regression
- Logistic Regression
- Decision trees
- Random forest
- Gradient boosting
- Support vector machine (SVM)
- Gradient descent
- Validation & cross-validation
- Over fitting & regularization
- Error decomposition

Module 5 : Unsupervised Learning Algorithms

- 
- Data dimensionality
 - Dimensionality reduction
 - Principal Component Analysis (PCA)
 - Kernel PCA
 - Centroid based clustering
 - K-means
 - K-medians
 - K-medoids
 - Density based clustering
 - DBSCAN
 - OPTICS
 - Distributing based clustering (Gaussian Mixture Model)

Final Module Ex

Part D Advanced Topics (50 Hs)

במודול זה נבססה נושאים בבחרים ומתקדמים מהעולמות המקיפים את מדען הנתונים.

- Intro to Deep learning
- Dimensionality reduction
- Neural networks
- Principal Component Analysis (PCA)
- Kernel PCA
- Intro to Recommendation Systems
- Intro to Anomaly Detection

Part E Practicum (40 Hs)