Intensive 16-week program curriculum

Week 01 - Introduction

- ➔ Overview of the curriculum
- ➔ Important roles to pursue in the future
  - ◆ Conventional Routes
    - Data Engineer
    - Machine Learning (ML) Engineer
    - Data Analyst
    - Data Scientist
    - Research Engineer
    - Research Scientist
    - Academic Researcher
  - ◆ Unconventional Routes
    - Freelance Data Science Engineer
    - Data Science Trainer or Mentor
    - Entrepreneur
- ➔ A day in the life (From Experience)
  - ◆ ML Engineer
  - ◆ Data Scientist
- ➔ Guide to Discipline (From Experience)
  - ◆ Limiting beliefs and how to fix them
  - ◆ Common reasons for failure
  - ◆ From Motivation to Discipline

Week 02 - Python Specialization

- ➔ Reviewing Basic Programming Concepts
  - ◆ Time Complexity
  - ◆ Sorting
  - ◆ Greedy Algorithms
  - ◆ Dynamic Programming
- ➔ Set Up Tools and Environments for Data Science
  - ◆ Virtual environments and why we need them
  - ◆ VS Code
  - ◆ Jupyter Notebook
  - ◆ Google Colaboratory
  - ◆ Kaggle
- ➔ How to Python
  - ◆ Script Structures
  - ◆ Conditionals and Loopings
  - ◆ Functions
  - ◆ Objects and Classes
  - ◆ Data Structures
  - ◆ Comprehensions
  - ◆ Debugging and Error Handling
- ➔ How to NumPy
  - ◆ Why use NumPy?
  - ◆ NumPy Basics
  - ◆ Array Manipulation

Week 03 - Data ETL (Extract, Transform, Load)

- ➔ Understanding End-to-end Data Science Pipeline
  - ◆ Data ETL
  - ◆ Data Analytics and Visualization
  - ◆ ML Modeling
  - ◆ Deployment and Performance Tracking
- ➔ Crawl a website
  - ◆ Why?
  - ◆ How? - Full Demonstration
- ➔ Using Pandas for ETL
  - ◆ Pandas Demo
    - Basics
    - Advanced Pandas Tricks
  - ◆ Data Manipulation
  - ◆ Quantitative Data Analysis

Week 04 - Data Visualization

- ➔ Visualization with Python
  - ◆ Matplotlib
  - ◆ Seaborn
  - ◆ Plotly
- ➔ Intro to Emerging Tools
  - ◆ Tableau
  - ◆ StreamLit

Week 05 - Data Analytics with Tableau

- ➔ Tableau Competitors
  - ◆ Microsoft Excel
  - ◆ KNIME
  - ◆ Power BI
- ➔ Why Tableau?
- ➔ Tableau Setup and ETL
- ➔ Filters
- ➔ Charts and Graphs
- ➔ Dashboards
- ➔ Parameters, Functions
- ➔ Blending, Joining

Week 06 - Data Analytics Capstone Project

A project with a particular problem statement will be assigned to judge the following skills

- ➔ Basic Web Scraping
- ➔ Data Manipulation
- ➔ Data Analysis
- ➔ Data Visualization
- ➔ Ability to use Tableau and StreamLit

Week 07 - Introduction to Deep Learning

- ➔ Origins to SOTA (State of the Art)
- ➔ Introducing necessary software
  - ◆ PyTorch
  - ◆ fastai
  - ◆ HuggingFace
  - ◆ Gradio
  - ◆ Weights and Biases
- ➔ Transfer Learning and Finetuning
- ➔ Train our first Computer Vision (CV) model
- ➔ Machine Learning (ML) and its types
- ➔ When to use ML
- ➔ How our CV model works
- ➔ What our CV model learns
- ➔ Train our first Segmentation, Tabular, and Recsys model
- ➔ Importance of Validation and Test Sets

Week 08 - End-to-end Data Science Pipeline in Practice

- ➔ Data Collection
- ➔ Data to DataLoaders
- ➔ Data Augmentation
- ➔ Training a model and Cleaning Data
- ➔ Deploy our first model
- ➔ Performance Monitoring
- ➔ Multi-label classification
- ➔ Regression

Week 09 - Image Recognizer Project and Data Ethics

- ➔ Individual Project: Build your own image recognizer with fastai from data gathering to model deployment
- ➔ Convolutional Neural Networks
- ➔ Residual Neural Networks
- ➔ Data Ethics
    - ◆ Examples of ethical disasters
    - ◆ The necessity of Data Ethics
    - ◆ Recourse and Accountability
    - ◆ Feedback Loops
    - ◆ Bias
    - ◆ Disinformation

Week 10 - Deep Dive into Deep Learning Foundations

- ➔ Building a baseline
- ➔ Gradient Descent
- ➔ Loss Functions
- ➔ Optimizer, Momentum, RMSProp, Adam
- ➔ Nonlinearities
- ➔ Building a PyTorch model from scratch
- ➔ Learning Rate Finder
- ➔ Discriminative Learning Rates
- ➔ Overfitting, extrapolation problems
- ➔ Exploding and vanishing gradient problems
- ➔ Propose team and project ideas for the final capstone project

Week 11 - Advanced Deep Learning Tricks

- ➔ Normalization
  - ◆ Batch Normalization
  - ◆ Layer Normalization
- ➔ Progressive Resizing
- ➔ Test Time Augmentation
- ➔ Mixup
- ➔ Label Smoothing
- ➔ Weight Decay
- ➔ Final capstone project discussion

Week 12 - Collaborative Filtering and Tabular Modeling

- ➔ Collaborative Filtering (Collab)
  - ◆ Where we need this
  - ◆ Processing the data
  - ◆ Collab model with PyTorch
  - ◆ Collab model with fastai
  - ◆ Interpreting the results
  - ◆ Custom Collab DL model
- ➔ Tabular Modeling
  - ◆ Processing the data
  - ◆ Decision Trees
  - ◆ Random Forests
  - ◆ Feature Importance
  - ◆ Data Leakage
- ➔ Final capstone project discussion

Week 13 - Natural Language Processing (NLP)

- ➔ Preprocessing Text
- ➔ Self-supervised learning
- ➔ RNN, LSTM
- ➔ Language Models with PyTorch and fastai
- ➔ Text Classifier with PyTorch and fastai
- ➔ Text Generation
- ➔ Risks of Language Models
- ➔ Transformers
- ➔ HuggingFace
- ➔ Final capstone project discussion

Week 14 - Experiment Management, Model Deployment, and Monitoring

- ➔ Experiment Management
  - ◆ Why do we need it?
  - ◆ EDA with Weights and Biases (W&B)
  - ◆ Project Management, Artifact Versioning
  - ◆ Collaboration tools
  - ◆ Hyperparameter sweeps
- ➔ Model Deployment
  - ◆ Batch Prediction
  - ◆ Rest APIs
  - ◆ Performance Optimization
  - ◆ Horizontal Scaling
  - ◆ Edge Prediction
- ➔ Model Monitoring
  - ◆ Why do we need it?
  - ◆ Domain Shift and Data Drift
  - ◆ What and How to Monitor
  - ◆ Tools for monitoring
- ➔ Final capstone project discussion

Week 15 and 16 - Deep Learning Capstone Project

A project of medium difficulty upon discussion with the course instructor. The project should include the following:

- ➔ Data Gathering, Processing, Cleaning
- ➔ Maintain the Data Ethics
- ➔ Building Models with PyTorch and/or fastai
- ➔ Utilizing some advanced deep learning tricks
- ➔ Experiment management with Weights and Biases
- ➔ Model deployment and monitoring