**Tóm Tắt Kiến Thức Đã Học**

Chương 1 – Giới thiệu về Machine Learning với Python

* + Machine Learning với Python nên tập trung vào thực hành, tránh học quá sa đà lý thuyết.
  + Mục tiêu: Xây dựng mô hình dự đoán ứng dụng được trong thực tế.
  + Các bước trong một dự án Machine Learning:
* Xác định bài toán
* Phân tích dữ liệu
* Tiền xử lý dữ liệu
* Đánh giá thuật toán
* Tối ưu kết quả
* Trình bày kết quả
  + Cuốn sách đi theo hướng "học qua dự án" thay vì lý thuyết nặng.

Chương 2 – Hệ Sinh Thái Python cho Machine Learning

* + Python là ngôn ngữ đa mục đích, dễ đọc, dễ học, lý tưởng cho Machine Learning.
  + SciPy ecosystem bao gồm:
    - NumPy: Làm việc với mảng dữ liệu.
    - Matplotlib: Trực quan hóa dữ liệu.
    - Pandas: Tổ chức và phân tích dữ liệu.
  + scikit-learn là thư viện Machine Learning chính, hỗ trợ nhiều thuật toán và công cụ.
  + Cách cài đặt môi trường:
    - Cài Python (nên dùng Python 3.8+).
    - Cài đặt SciPy stack: scipy, numpy, matplotlib, pandas.
    - Cài scikit-learn.
    - Hoặc dùng Anaconda để cài đặt nhanh tất cả.
  + Lợi ích: Có thể làm R&D và triển khai production trên cùng một nền tảng Python.
* Chương 3 – Khóa Học Cấp Tốc Python và SciPy
  + Ôn lại các cấu trúc dữ liệu cơ bản trong Python: Tuple, List, Dictionary, Hàm.
  + Hiểu rõ cú pháp điều kiện (if-else), vòng lặp for và while.
  + Làm việc với NumPy:
    - Tạo mảng, truy cập phần tử, thực hiện phép toán vector.
  + Dùng Matplotlib để trực quan hoá dữ liệu với biểu đồ đường và biểu đồ điểm.
  + Dùng Pandas để tạo Series và DataFrame:
    - Truy xuất dữ liệu bằng chỉ số, tên cột.
    - In dữ liệu và hiểu cách tổ chức dạng bảng.
* Chương 4 – Cách Nạp Dữ Liệu Machine Learning
  + Hiểu định dạng CSV là phổ biến trong Machine Learning.
  + Biết các vấn đề thường gặp khi load CSV: header, comment, delimiter, quote.
  + Làm quen với tập dữ liệu Pima Indians – dự đoán tiểu đường.
  + Ba cách nạp dữ liệu CSV:
* Sử dụng thư viện chuẩn csv → cần xử lý thô, đổi sang NumPy.
* Sử dụng NumPy → dùng loadtxt, nhanh, gọn.
* Sử dụng Pandas → linh hoạt, dễ gán tên cột, dễ khám phá.
  + Kết luận: nên dùng Pandas cho hầu hết dự án thực tế.