Machine Learning with Python

INTRODUCTION TO MACHINE LEARNING & BASICS OF PYTHON

Module 1: OVERVIEW OF MACHINE LEARNING

ML Spectrum & Journey

Intro to Modelling Lifecycle

Intro to Supervised Learning

Intro to Unsupervised Learning

Module 2: INTRODUCTION TO PYTHON

Python Basics

Spyder IDE

Jupyter Notebook

Floats and Strings

Simple Input & Output

Variables

Single and Multiline Comments

Module 3: CONTROL STRUCTURES

Booleans and Comparisons

Conditional Statements (IF ELSE) Operator Precedence

Lists - Operations and Functions

Module 4: FUNCTIONS AND MODULES

Function Arguments

Comments and doc strings

Functions as Objects

Modules

Standard Lib and Pip

Module 5: EXCEPTIONS AND FILES

Exception Handling

Raising Exceptions

Assertions

Working with Files

INTRODUCTION TO STATISTICS

Module 1: Basic Probability

Rules of Probability

Permutations and Combinations

Bayes Theorem

Descriptive Statistics

Compound Probability

Conditional Probability

Module 2: Probability Distributions

Types of Distributions

Functions of Random Variables

Probability Distribution Graphs

Confidence Intervals

Module 3: Data Transformation

Merge, Rollup, Transpose and Append

Missing Analysis and Treatment

Outlier Analysis and Treatment

Module 4: Exploratory Data Analysis

Summarizing and Visualizing the Important Characteristics of Data

Hypothesis Testing

Visualizations

Univariates, Bi-variates

Crosstabs, Correlation

DATA ANALYSIS FOR MACHINE LEARNING

Module 1: PANDAS

Introduction to Pandas

IO Tools

Basics of NumPy

NumPy Functions

Pandas - Series and Dataframes

Module 2: Data Visualisation

Basics of Data Visualization

Line Plots

Bar Charts

Pie Charts

Histograms

Scatter Plots

Parallel Coordinates

Project

Module 3: Linear Regression

Implementing Simple & Multiple Linear Regression with Python

Making Sense of Result Parameters

Model Validation

Handling other Issues/Assumptions in Linear Regression: Handling Outliers, Categorical Variables, Autocorrelation, Multicollinearity, Heteroskedasticity

Prediction and Confidence Intervals

Project

Module 4: Logistics Regression

Logistic Regression with Python

Making Sense of Result

Parameters: Wald Test, Likelihood Ratio Test Statistic, Chi-Square Test

Goodness of Fit Measures

Model Validation: Cross Validation, ROC Curve, Confusion Matrix

Project

Module 5: Logistics Regression

Implementing Decision Trees using Python

Homogeneity

Entropy

Information Gain

Gini Index

Standard Deviation Reduction

Vizualising & Prunning a Tree

Project

MACHINE LEARNING

Module 1: INTRODUCTION TO MACHINE LEARNING

Machine Learning

ML Modelling Flow

How to Treat Data in ML

Parametric & Non-Parametric ML Algorithm

Types of Machine Learning

Performance Measures

Bias-Variance Trade-Off

Overfitting & Underfitting

Bootstrap Sampling

Bagging Aggregation

Boosting

Module 2: OPTIMISATION TECHNIQUES

Constant Learning Rate Procedures

Adaptive Learning Procedures

Batch Gradient Descent

Mini-Batch Gradient Descent

Stochastic Gradient Descent

Nesterov Accelerated Gradient

Root Mean Squared Propagation

Adaptive Moment Estimation Procedure

Module 3: ML ALGORITHM - SUPERVISED LEARNING AND UNSUPERVISED LEARNING

Linear Regression with Stochastic Gradient Descent

Logistic Regression with Stochastic Gradient Descent

K-Nearest Neighbour

Eager Methods vs. Lazy Methods

Nearest Neighbor Classification

Building kD-Trees

Support Vector Machine

Perceptron Algorithm. What is Clustering?

K-means Algorithm

K-means Clusters

Module 4: ENSEMBLE ALGORITHMS

Ensemble Techniques

Bootstrap Aggregation

Random Forest

Boosting

PROJECTS

Projects are some of the best investments of your time. You’ll enjoy learning, stay motivated, and make faster progress.

You see, no amount of theory can replace hands-on practice. Textbooks and lessons can lull you into a false belief of mastery because the material is there in front of you. But once you try to apply it, you might find that it’s harder than it looks.

Projects help you improve your applied ML skills quickly while giving you the chance to explore an interesting topic.

Plus, you can add projects into your portfolio, making it easier to land a job, find cool career opportunities, and even negotiate a higher salary.

1. Machine Learning Sparta

We’re affectionately calling this “machine learning sparta,” but it’s not new. This is one of the fastest ways to build practical intuition around machine learning.

The goal is to take out-of-the-box models and apply them to different datasets. This project is awesome for 3 main reasons:

First, you’ll build intuition for model-to-problem fit. Which models are robust to missing data? Which models handle categorical features well? Yes, you can dig through textbooks to find the answers, but you’ll learn better by seeing it in action.

Second, this project will teach you the invaluable skill of prototyping models quickly. In the real world, it’s often difficult to know which model will perform best without simply trying them.

Finally, this exercise helps you master the workflow of model building. For example, you’ll get to practice…

* Importing data
* Cleaning data
* Splitting it into train/test or cross-validation sets
* Pre-processing
* Transformations
* Feature engineering

Because you’ll use out-of-the-box models, you’ll have the chance to focus on honing these critical steps.

2. Play Money Ball

In the book [Moneyball](https://en.wikipedia.org/wiki/Moneyball), the Oakland A’s revolutionized baseball through analytical player scouting. They built a competitive squad while spending only 1/3 of what large market teams like the Yankees were paying for salaries.

In project you’ll learn

* Sports betting… Predict box scores given the data available at the time right before each new game.
* Talent scouting… Use college statistics to predict which players would have the best professional careers.
* General managing…Create clusters of players based on their strengths in order to build a well-rounded team.

Sports is also an excellent domain for practicing data visualization and exploratory analysis. You can use these skills to help you decide which types of data to include in your analyses.

3. Predict Stock Prices

The stock market is like candy-land for any data scientists who are even remotely interested in finance.

First, you have many types of data that you can choose from. You can find prices, fundamentals, global macroeconomic indicators, volatility indices, etc… the list goes on and on.

Second, the data can be very granular. You can easily get time series data by day (or even minute) for each company, which allows you think creatively about trading strategies.

Finally, the financial markets generally have short feedback cycles. Therefore, you can quickly validate your predictions on new data.

Some examples of beginner-friendly machine learning projects you could try include…

* Quantitative value investing… Predict 6-month price movements based fundamental indicators from companies’ quarterly reports.
* Forecasting… Build time series models, or even recurrent neural networks, on the delta between implied and actual volatility.
* Statistical arbitrage… Find similar stocks based on their price movements and other factors and look for periods when their prices diverge.

Obvious disclaimer: Building trading models to practice machine learning is simple. Making them profitable is extremely difficult. Nothing here is financial advice, and we do not recommend trading real money.

4. Investigate Enron

The [Enron scandal and collapse](http://www.investopedia.com/updates/enron-scandal-summary/) was one of the largest corporate meltdowns in history.

In the year 2000, Enron was one of the largest energy companies in America. Then, after being outed for fraud, it spiraled downward into bankruptcy within a year.

Luckily for us, we have the Enron email database. It contains 500 thousand emails between 150 former Enron employees, mostly senior executives. It’s also the only large public database of real emails, which makes it more valuable.

In fact, data scientists have been using this dataset for education and research for years.

Examples of machine learning projects for beginners you will try are

* Anomaly detection… Map the distribution of emails sent and received by hour and try to detect abnormal behaviour leading up to the public scandal.
* Social network analysis… Build network graph models between employees to find key influencers.

## 5. Write ML Algorithms from Scratch

Writing machine learning algorithms from scratch is an excellent learning tool for two main reasons.

First, there’s no better way to build true understanding of their mechanics. You’ll be forced to think about every step, and this leads to true mastery.

Second, you’ll learn how to translate mathematical instructions into working code. You’ll need this skill when adapting algorithms from academic research.

To start, we recommend picking an algorithm that isn’t too complex. There are dozens of subtle decisions you’ll need to make for even the simplest algorithms.

After you’re comfortable building simple algorithms, try extending them for more functionality. For example, try extending a vanilla **logistic regression** algorithm into a **lasso/ridge regression** by adding regularization parameters.

Finally, here’s a tip every beginner should know: Don’t be discouraged if your algorithm is not as fast or fancy as those in existing packages. Those packages are the fruits of years of development!

6. Mine Social Media Sentiment

Social media has almost become synonymous with “big data” due to the sheer amount of user-generated content.

Mining this rich data can prove unprecedented ways to keep a pulse on opinions, trends, and public sentiment. Facebook, Twitter, YouTube, WeChat, WhatsApp, Reddit… [the list goes on and on](https://www.fool.com/investing/2017/03/30/top-10-social-networks-how-many-users-are-on-each.aspx).

Furthermore, every generation is spending even more time on social media than their predecessors. This means that social media data is will become even more relevant for marketing, branding, and business as a whole.

While there are many popular social media platforms out there, Twitter is the classic entry point for practicing machine learning.

With Twitter data, you get an interesting blend of data (tweet contents) and meta-data (location, hashtags, users, re-tweets, etc.) that open up nearly endless paths for analysis.

7. Improve Health Care

Another industry that’s undergoing rapid changes thanks to machine learning is global health and health care.

In most countries, becoming a doctor requires many years of education. It’s a demanding field with long hours, high stakes, and an even higher barrier to entry.

As a result, there has recently been significant effort to alleviate doctors’ workload and improve the overall efficiency of the health care system with the help of machine learning.

Projects include

* **Preventative care**… Predicting disease outbreaks on both the individual and the community level.
* **Diagnostic care**… Automatically classifying image data, such as scans, x-rays, etc.
* **Insurance**… Adjusting insurance premiums based on publicly available risk factors.

As hospitals continue to modernize patient records and as we collect more granular health data, there will be an influx of low-hanging fruit opportunities for data scientists to make a difference.