Machine Learning in Python

Inside this lesson, you will learn how to perform machine learning in Python on numerical data and image data.

You will learn how to operate popular Python machine learning and deep learning libraries, including two of my favorites:

- scikit-learn
- Keras

Specifically, **you will learn how to**:

1. Examine your problem
2. Prepare your data (raw data, feature extraction, feature engineering, etc.)
3. Spot-check a set of algorithms
4. Examine your results
5. Double-down on the algorithms that worked best

Using this technique you will be able to get your start with machine learning and Python!

Along the way, **you'll discover popular machine learning algorithms that you can use in your own projects as well**, **including**:

1. k-Nearest Neighbors (k-NN)
2. Naïve Bayes
3. Logistic Regression
4. Support Vector Machines (SVMs)
5. Decision Trees
6. Random Forests
7. Perceptrons

8. Multi-layer, feedforward neural networks

9. Convolutional Neural Networks (CNNs)

This hands-on experience will give you the knowledge (and confidence) you need to apply machine learning in Python to your own projects.

Lets look at data science pipeline

7 **Fundamental Steps to Complete a Data Analytics Project**
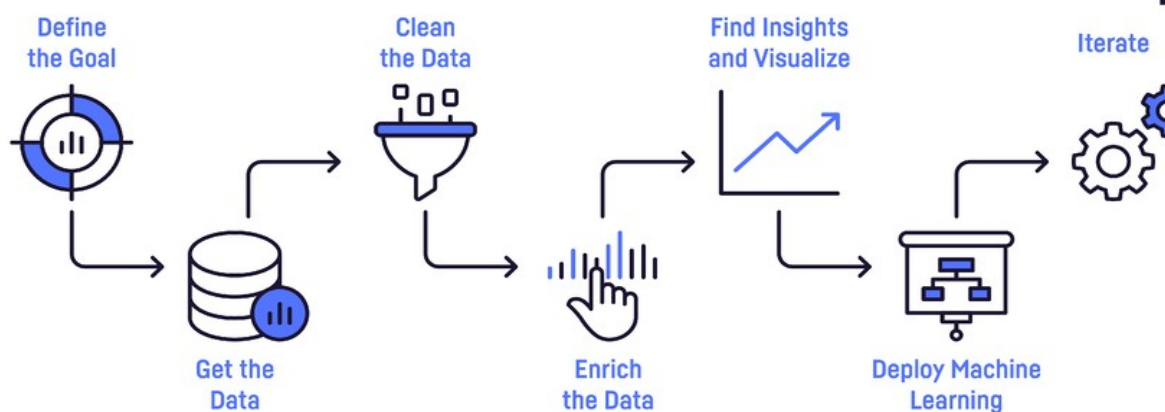
Step 1: **Understand the Business**

Understanding the business or activity that your data project is part of is key to ensuring its success and the first phase of any sound data analytics project.

If you're working on a personal project or playing around with a dataset (Learning), this step may seem irrelevant. Buts It's not. Simply downloading a cool open dataset is not enough. In order to have motivation, direction, and purpose, you have to

identify a clear objective of what you want to do with data: a concrete question to answer, a product to build, etc.

Step 2: Get Your Data

Once you've gotten your goal figured out, it's time to start looking for your data, the second phase of a data analytics project. Mixing and merging data from as many data sources as possible is what makes a data project great

Here are a few ways to get some usable data:

Connect to a database: Ask your data and IT teams for the data that's available or open up your private database and start accessing through it to understand the data set.

Use APIs: Think of the APIs to all the tools your company's been using and the data that have been collected.

Look for open data: The Internet is full of datasets to enrich what you have with additional information. For example, census data will help you add the average revenue for the district where your user lives or OpenStreetMap can show you how many coffee shops are on a given street. A lot of countries have open data platforms I.e

https://www.opendata.go.ke/ in kenya,

like data.gov in the U.S,

https://archive.ics.uci.edu/ml/datasets.php ,

https://www.kaggle.com/

and many more . .

Step 3: **Explore and Clean Your Data**

The next data science step is the dreaded data preparation process that typically takes up to 80% of the time dedicated to a data project.

Once you've gotten your data, it's time to get to work on it in the third data analytics project phase. Start looking at the data to see what you've got and how you can link everything together to achieve your original goal. Start taking notes on your first analyses and ask questions to business people, the IT team, or other groups to understand what all your variables mean.

The next step (and by far the most dreaded one) is cleaning your data. You've probably noticed that even though you have a country feature, for instance, you've got different spellings, or even missing data. It's time to look at every one of your columns to make sure your data is homogeneous and clean.

Step 4: **Enrich Your Dataset**

Now that you have clean data, it's time to manipulate it in order to get the most value out of it. You should start the data enrichment phase of the project by joining all your different sources and group logs to narrow your data down to the essential features.

Step 5: **Build Helpful Visualizations**

You now have a nice dataset (or maybe several), start exploring it by building graphs. When you're dealing with large volumes of data, visualization is the best way to explore and communicate your findings and is the next phase of your data analytics project.

Use **seaborn, plotly, Power BI, tableau** and any other to create your plots.

Step 6: **Modeling** - **Deploy Machine Learning Models**

The next data science step, phase six of the data project, is when the modeling start. Machine learning algorithms can help you go a step further into getting insights and predicting future trends.

By working with supervised or unsupervised learning, you can build models to uncover trends in the data that were not distinguishable in graphs and stats. Here you can create predictions based on model training, or create groups of similar events (or clusters) and more or less explicitly express what feature is decisive in these results.

Step 7: **Communicate** & **Iterate**, **Iterate**, **Iterate**

The main goal in any business project is to prove its effectiveness as fast as possible to justify, well, your job. The same goes for data projects.

Here communicate your results and continue improving your model by revisiting earlier steps, this helps improve model overtime, Note data is dynamic!

According to O'Reilly, one of the biggest mistakes that people make with regard to machine learning is thinking that once a model is built and goes live, it will continue working as normal indefinitely. On the contrary, models will actually degrade in quality over time if they're not continuously improved and fed new data.

Check this notebook for your first machine learning project!

https://colab.research.google.com/drive/1SuakVSGSbhZoj2Fx1iiZv8a24LyRw7IR?usp=sharing

More Class Examples