

◆ Member-only story

A Simple Example of Pipeline in Machine Learning with Scikit-learn



Saptashwa Bhattacharyya · Follow

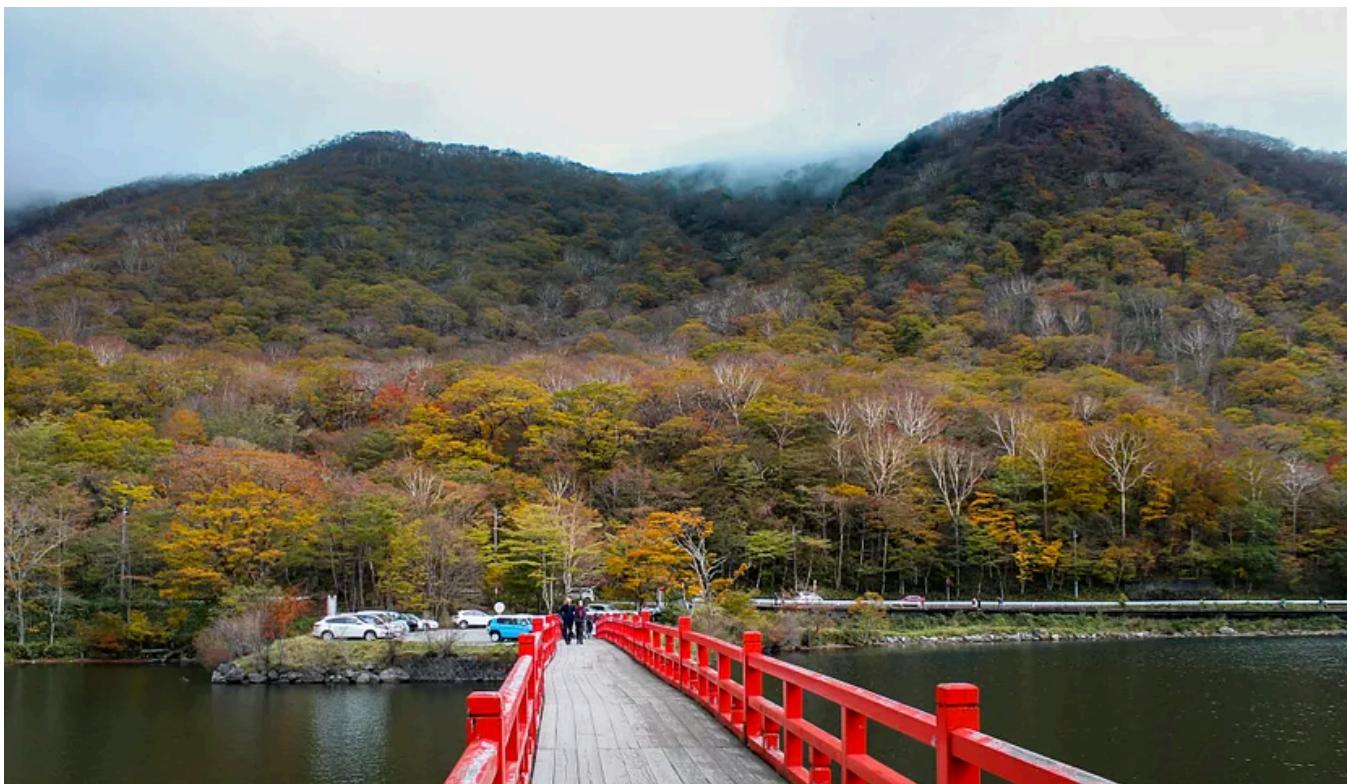
Published in Towards Data Science

5 min read · Nov 12, 2018

Listen

Share

More



Open in app ↗

Medium



Search



Today's post will be short and crisp and I will walk you through an example of using Pipeline in machine learning with python. I will use some other important tools like GridSearchCV etc., to demonstrate the implementation of pipeline and finally explain why pipeline is indeed necessary in some cases. Let's begin

Definition of pipeline class according to scikit-learn is

Sequentially apply a list of transforms and a final estimator. Intermediate steps of pipeline must implement fit and transform methods and the final estimator only needs to implement fit.

The above statements will be more meaningful once we start to implement pipeline on a simple data-set. Here I'm using the red-wine data-set, where the 'label' is quality of the wine, ranging from 0 to 10. In terms of data pre-processing, it's a rather simple data-set as, it has no missing values.

Yavorski, read the best stories about data science from industry leaders on Medium.

The author made this story available to Medium members only. Upgrade to instantly unlock this story plus other member-only benefits.

- ◆ Access all member-only stories on Medium
- ◆ Become an expert in your areas of interest
- ◆ Get in-depth answers to thousands of questions about data science
- ◆ Grow your career or build a new one



Marc-André Giroux
Sr. Software Developer



Carlos Arguelles
Sr. Staff Engineer



Tony Yiu
Director



Brandeis Marshall
CEO

Netflix

Google

Nasdaq

DataedX



Cassie Kozyrkov
Chief Decision Scientist
Google



Memo Akten
Asst. Professor
UCSD



Vitali Zaidman
Software Architect
Meta



Camille Fournier
Head of Engineering
JPMorgan Chase

Resume membership



Follow



Written by Saptashwa Bhattacharyya

3.6K Followers · Writer for Towards Data Science

PhD, Astrophysics. Using Deep Learning, Searching Dark Matter! <https://www.linkedin.com/in/saptashwa>

More from Saptashwa Bhattacharyya and Towards Data Science

$$m_1^{(t)} = \beta_1 m_1^{(t-1)} + (1 - \beta_1) g^{(t)}, \hat{m}_1^{(t)} = \frac{m_1^{(t)}}{1 - \beta_1^t}$$

$$m_2^{(t)} = \beta_3 m_2^{(t-1)} + (1 - \beta_3) g^{(t)}$$

$$v^{(t)} = \beta_2 m_1^{(t-1)} + (1 - \beta_2) g^{(t)}, \hat{v}^{(t)} = \frac{v^{(t)}}{1 - \beta_2^t},$$

$$\theta^{(t)} = \theta^{(t-1)} - \eta \left(\frac{\hat{m}_1^{(t)} + \alpha \hat{m}_2^{(t)}}{\sqrt{\hat{v}^{(t)}} + \epsilon} + \lambda \theta^{(t-1)} \right)$$



Saptashwa Bhattacharyya in Towards Data Science

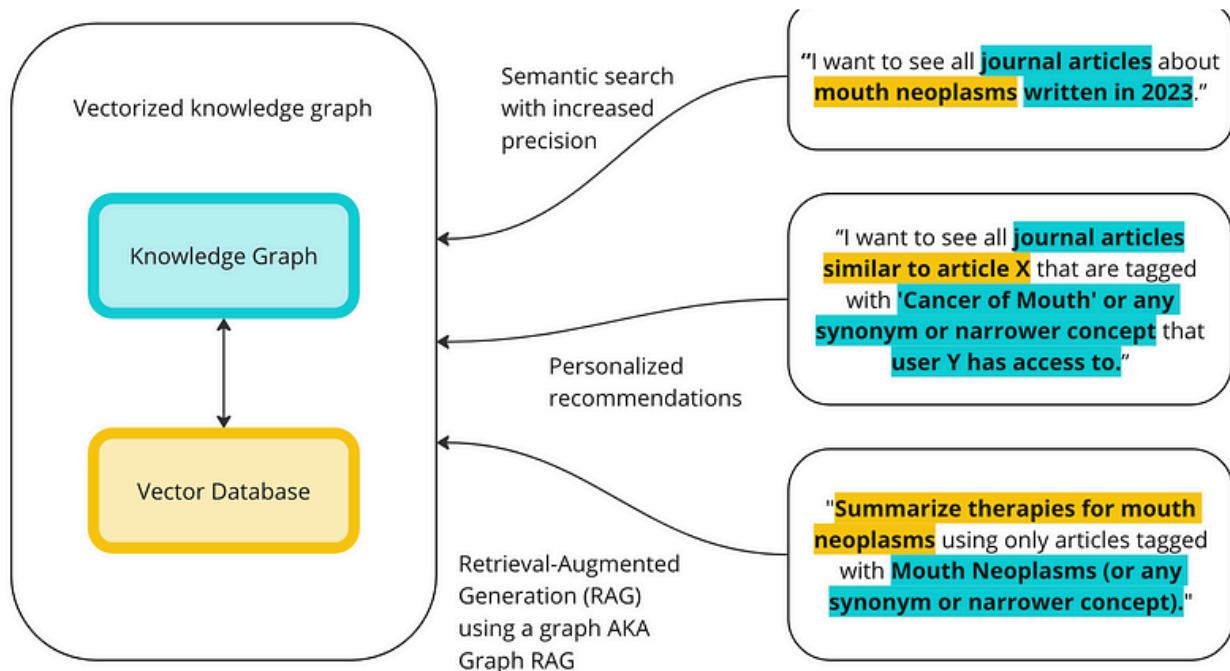
AdEMAMix: A Deep Dive into a New Optimizer for Your Deep Neural Network

A better and faster option than the ADAM optimizer, from Apple Research

4 Sep 19 16



...



Steve Hedden in Towards Data Science

How to Implement Graph RAG Using Knowledge Graphs and Vector Databases

A Step-by-Step Tutorial on Implementing Retrieval-Augmented Generation (RAG), Semantic Search, and Recommendations

Sep 6 1.2K 13



 Ahmed Besbes in Towards Data Science

What Nobody Tells You About RAGs

A deep dive into why RAG doesn't always work as expected: an overview of the business value, the data, and the technology behind it.

Aug 23 1.7K 24



$$\Sigma_l = \begin{bmatrix} \phi_{1,1}(\cdot) & \phi_{1,2}(\cdot) & \dots & \phi_{1,n_l}(\cdot) \\ \phi_{2,1}(\cdot) & \phi_{2,2}(\cdot) & \dots & \phi_{2,n_l}(\cdot) \\ \vdots & \ddots & \ddots & \vdots \\ \phi_{n_{l+1},1} & \dots & & \phi_{n_{l+1},r} \end{bmatrix}$$

 Saptashwa Bhattacharyya in Towards Data Science

KAN: Why and How Does It Work? A Deep Dive

Can we discover new physics with KAN?

★ May 10 ⌘ 451 🗣 3

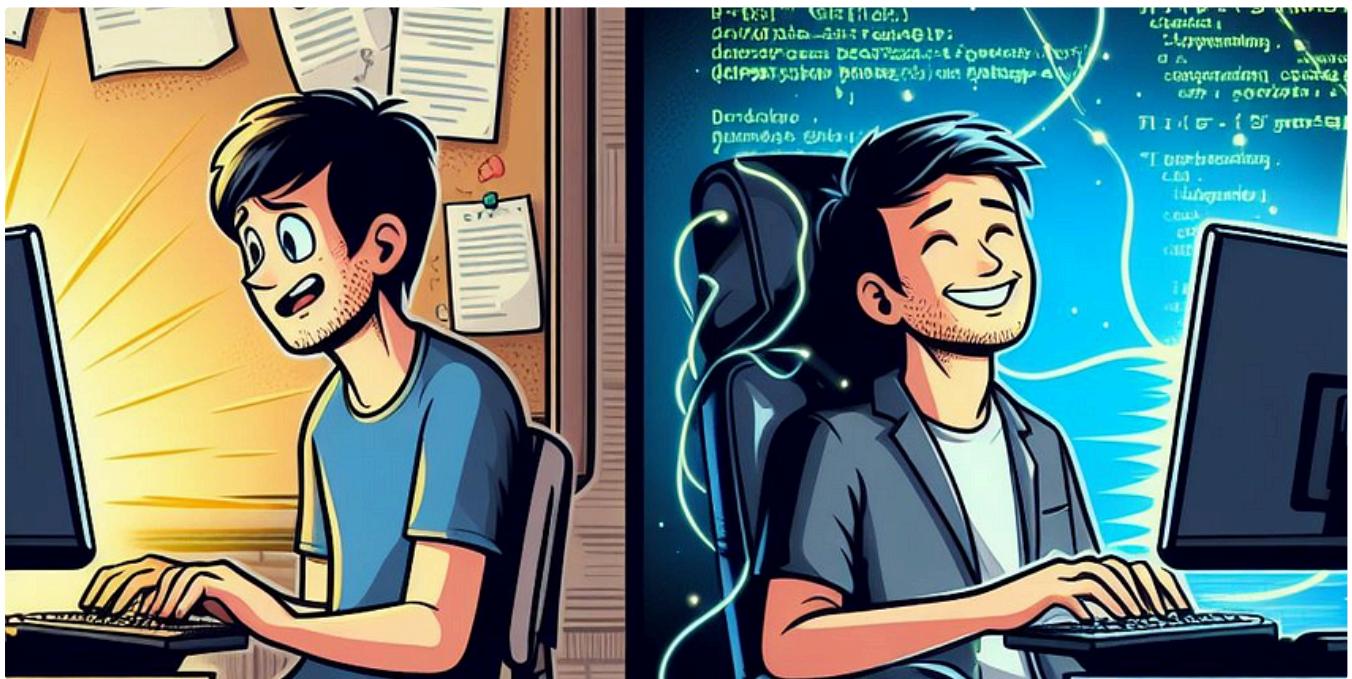


...

See all from Saptashwa Bhattacharyya

See all from Towards Data Science

Recommended from Medium

 Abhay Parashar in The Pythoneers

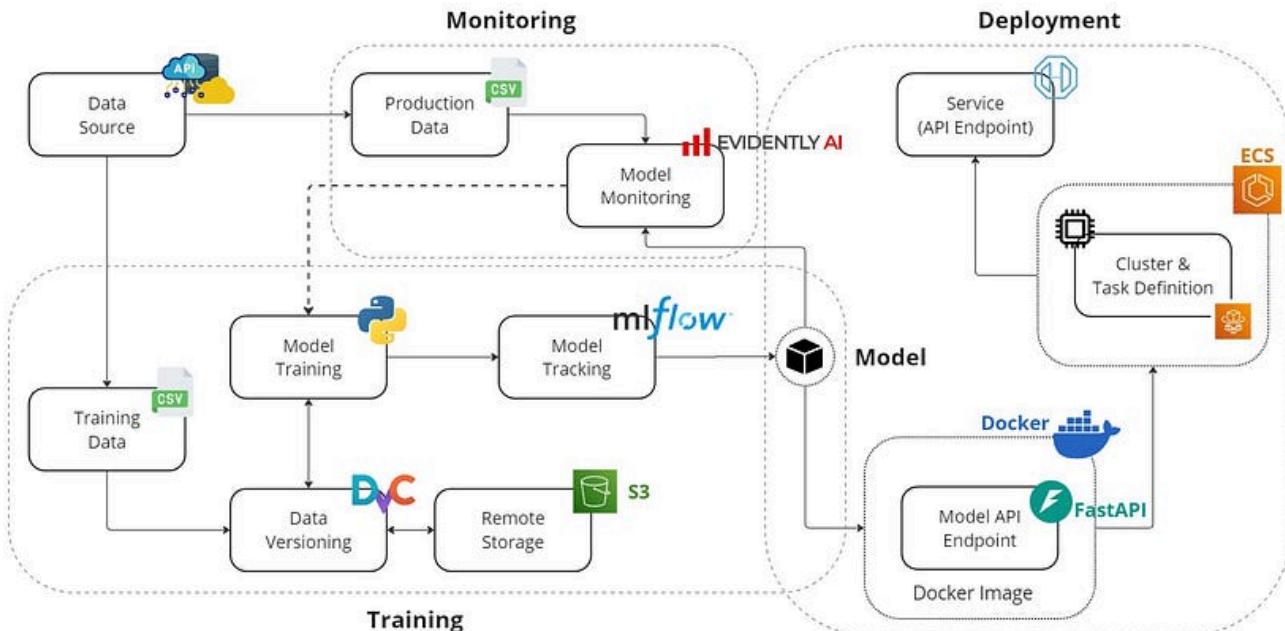
17 Mindblowing Python Automation Scripts I Use Everyday

Scripts That Increased My Productivity and Performance

★ Aug 25 ⌘ 7.9K 🗣 75



...



Prasad Mahamulkar

Machine Learning Operations (MLOps) For Beginners

End-to-end Project Implementation

Aug 29 · 1.1K · 8



Lists



Predictive Modeling w/ Python

20 stories · 1558 saves



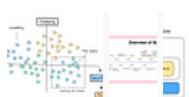
Practical Guides to Machine Learning

10 stories · 1889 saves



Coding & Development

11 stories · 821 saves



Natural Language Processing

1728 stories · 1299 saves



Bhavik Jikadara

Getting Started with MLOps: A Beginner's Guide

The Comprehensive Guide to MLOps

Mar 31 107



...

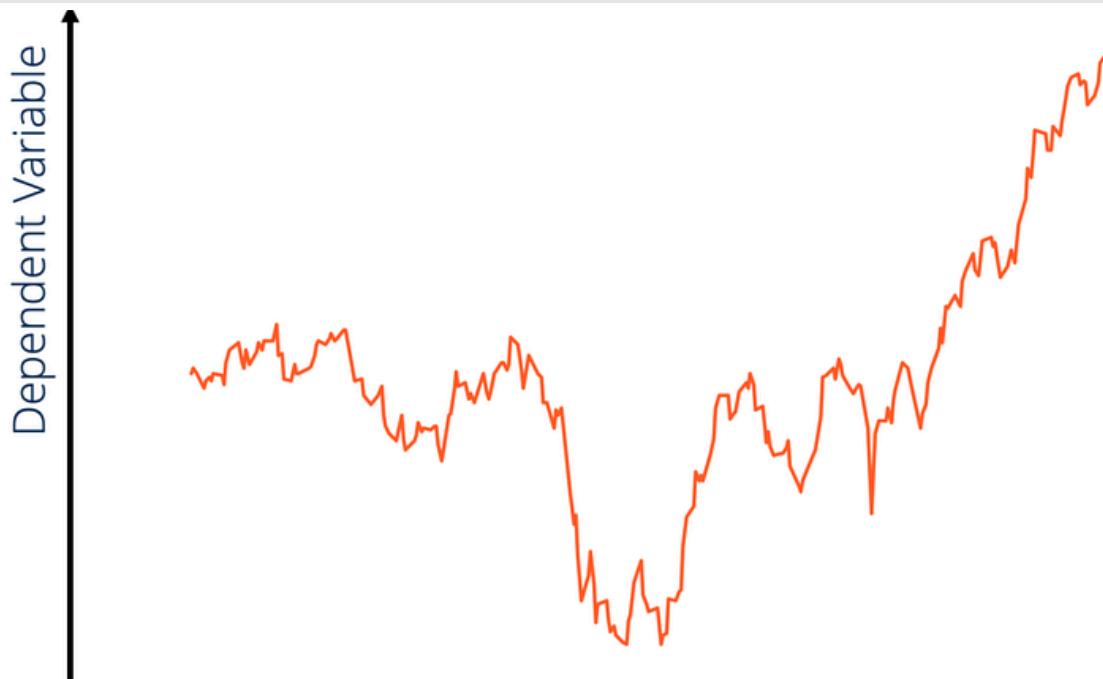


Lazar Gugleta in Towards AI

Why Polars Destroys Pandas in All Possible Ways for Data Scientists?

One of the first Data Science libraries, Pandas, has been improving the lives of many developers across the globe, but Polars shows it is...

Aug 8 781 12



Palash Mishra

Predicting Time Series Data with Machine Learning, Generative AI, and Deep Learning

Time series data prediction is a critical aspect of various industries, ranging from finance and healthcare to marketing and logistics. The...

Jun 19 703 9




`sys.getsizeof([0] * 3) → 80`
`sys.getsizeof([0, 0, 0]) → 120`
`sys.getsizeof([0 for i in range(3)]) → 88`

 Shuai Li in Programming Domain

Most Developers Failed with this Senior-Level Python Interview Question

What is the difference between `[0] * 3` and `[0, 0, 0]`?

Aug 9

3.1K

47



...

See more recommendations