# https://lh4.googleusercontent.com/YAkJn7VqqmAS8Vv8-JktR7RmdchyEDaNQUZSWUg3c1E1rKm9PSRoZwMP1pBLNrGNZ_sdrzO83F1Kg_RJHQis9STLyEw8TN12v56np9ZB7Qmwb_WAsfmGdxKFLcSFMUnIe-5ZJczX

# **Hands-On Machine Learning with Python and Scikit Learn**

***Develop and apply the best Machine Learning practices using the powerful features of Python and Scikit Learn.***

# **About the author:**

*Taylor smith is a machine learning enthusiast with over five years of experience who loves to apply interesting computational solutions to challenging business problems. Currently working as Principal Data-scientist, Taylor is also an active open-source contributor and staunch pythonista.*[*LinkedIn*](https://www.linkedin.com/in/taylorgsmith/) *and* [*Github*](https://github.com/tgsmith61591) *profile.*

# **OVERVIEW**

Machine learning and “artificial intelligence” are the new “big data”—at least as far as buzzwords in the workplace go. The scikit-learn library is one of the most popular platforms for everyday machine learning and data science. The reason is because it is built upon Python, a fully featured programming language. This course will help to discover the magical “black box” that is machine learning by teaching a practical approach to modeling using Python along with Scikit Learn library.

We’ll begin the journey by quickly wrapping up some of the exciting and important machine learning practices before moving into exploratory data analysis. Moving along, we’ll learn to develop complex pipelines and techniques for building custom transformer objects for feature extraction, manipulation, and other effective techniques of data cleansing. Finally, we’ll discover how to select a model, tune its optimal hyper-parameters and deploy.

This video course highlights clean coding techniques, object-oriented transformer design and the best practices in machine learning while using the Scikit Learn library and also maintaining focus on practicality and re-usability ensuring this can be applied throughout machine learning projects of all size.

# **TARGET AUDIENCE**

This course is aimed at students, data-scientists with prior Python programming experience and looking to upgrade their machine learning skills using Python.

Basic familiarity or exposure to some level of statistics is recommended, but not required.

# 

# **APPROACH**

1. The course will be a practical and step-by-step guide to implement effective machine learning techniques using Scikit library. Exercises are provided to reinforce key topics and support learning.

2) An easy-to-follow implementation of scikit-learn library that will help you to get started with the effective machine learning techniques using Python.

# **FEATURES**

* Deep dive into machine learning using the most advanced tools and Scikit library.
* Develop complex pipelines and processing of data through manipulation, extraction and data-cleansing techniques.
* Clean coding techniques and best practices in machine learning which can be applicable to any scalable machine learning projects.

# 

# **WHAT WILL YOU LEARN**

· Split data effectively using the Scikit Machine learning model

· Explore, organize, manipulate and analyze your data (including some visual & descriptive statistics techniques)

· Enhance your model performance using Cross validation

· Build and design your model by developing a baseline model and pipelining transformers

· Selecting an optimal model and tuning hyper-parameters so that the model can best solve the machine learning problem

· Persisting a model for use in production

# **COURSE OUTLINE (3.5 hrs)**

**Section 1: Getting Started with Machine Learning (45 min)**  
This volume starts explaining how to set up our environment to build a machine learning application by leveraging a toolkit of open source libraries. First, we’ll discuss proper setup our Anaconda environment and from there, we will dive deeper into exploring a simple dataset (preferably IRIS), some important ML concepts and familiarizing ourselves with our tools.  
  
• Video 1.1: Demo machine learning product (5 mins)  
• Video 1.2: Setting up our Anaconda environment (5 min)  
• Video 1.3: Introduction to the iPython notebook (3 min)   
• Video 1.4: Loading and manipulating data with Pandas (5 min)   
• Video 1.5: ML objective + data splitting and common pitfalls (15 min)  
• Video 1.6: Descriptive statistics using Pandas (5 min)  
• Video 1.7: Planning our preprocessing stages (7 min)

* Assessment

**Section 2: Exploratory data analysis (39 min)**  
It’s time to start exploring our dataset! In this section, we’ll discuss proper preparation of our dataset, how best to handle messy data, creating new, informative variables, and selecting features to use for modeling. In addition, we’ll explore some descriptive analytics and discuss some of the common pitfalls of the exploratory data analysis phase.  
  
• Video 2.1: Imputing missing values (7 min)  
• Video 2.2: Handling categorical data (10 min)   
• Video 2.3: Handling outliers (5 min)  
• Video 2.4: Feature extraction/engineering & custom transformers (10 min)  
• Video 2.5: Feature selection (7 min)

* Assessment

**Section 3: Building your first model (51 min)**  
Now that we’ve prepared our dataset, we can build our first model! This section will cover pipelining our transformers and baselining performance. Moreover, we’ll discuss some practical techniques using Scikit library that will bolster your model performance and improve test set generalizability.  
  
• Video 3.1: Pipelining transformers (7 min)  
• Video 3.2: Bias/variance trade-off, overfitting and underfitting (10 min)  
• Video 3.3: Cross validation (12 min)  
• Video 3.4: Scoring metrics (12 min)  
• Video 3.5: Developing model baselines (10 min)

* Assessment

**Section 4: Model tuning & selection using Scikit Learn (33 min)**  
Now that we have a baseline model, we want to squeeze all the performance out of it that we can. This section will cover techniques related to selecting optimal model “hyper-parameters” in a way that minimizes the likelihood of over-fitting our data.  
  
• Video 4.1: Hyper-parameters & strategic search ranges (10 min)  
• Video 4.2: The importance of cross validation (7 min)  
• Video 4.3: Model wars using grid searches (10 min)   
• Video 4.4: Final model selection, and exposure to holdout set (6 min)

* Assessment

**Section 5: Model deployment (30 min)**  
Having selected a model is not enough! This section will address common misconceptions over what happens next, best practices around model deployment, and practical solutions for deploying a simple model via a RESTful API.  
  
• Video 5.1: Model selection—where do we go now? (5 min)  
• Video 5.2: Model persistence and loading to/from disk (7 min)  
• Video 5.3: A brief note on persistence and version perils—also conda envs in deployment (6 min)   
• Video 5.4: Deployment ML applications behind a RESTful endpoint using Flask (12 min)

* Assessment