

AI & ML for Data Scientists

Class 1: Why AI and Machine Learning for Economists

Ge Lei

Renmin University of China – Quantitative Economics

September 24, 2025



中國人民大學
RENMIN UNIVERSITY OF CHINA

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

Teaching Team (Finance)

- Ge Lei (Quantitative Economics, Renmin University)
 - Dr. Li Fengxin (School of Information, Renmin University)
 - Liang Hongming (Experimental Class of Economics and Mathematics, Renmin University)
 - Ouyang Yubo (Experimental Class of Economics and Mathematics, Renmin University)

Teaching Team (Econometrics)

- Ge Lei (Quantitative Economics, Renmin University)
- Wang Chenxi (Experimental Class of Finance and Mathematics, Renmin University)
- Cui Tianyou (Experimental Class of Finance and Mathematics, Renmin University)

Teaching Team (English)

- Ge Lei (Quantitative Economics, Renmin University)
 - Liang Wei (Experimental Class of Finance and Mathematics, Renmin University)
 - Xue Jiayi (School of Economics, Renmin University)

Why Economic Scientists Should Engage in Quantitative Modeling?

- A future essential skill for economists
- Machine-learning algorithms vastly outperform human analysts

Example: AI Accuracy and Speed Outperform Real-Estate Analysts

- Prediction error for distressed real-estate asset packages is only **3%**
- Accuracy for predicting single property values is about 5% (analyst error 8%)
- Evaluate tens of thousands of properties per hour (analyst 1 property / hour)

Expected Outcomes

1. Critical Thinking (model-based reasoning and logic)
2. Ability to handle financial and economic data with Python (Data Scientist)
3. Ability to build quantitative models for analysis and prediction with Python

Prerequisites Before Taking the Course

1. Economics: all of you have undergone 3 years of rigorous training.
2. Mathematics: 6 years of intensive training from high school onwards.
3. English: 9 years of rigorous training from junior high.

Bottom line: Quantitative modeling = Economics + Mathematics + English + Python. It's the stage where years of training can finally be showcased.

Course Assessment

The course assessment simulates the real work of a company's quantitative department, training programming, teamwork, free-rider issues, practical business projects, and economic presentation skills – all relevant for academics as well.

- Homework 16%: group programming assignments (No Free Rider) – 5 assignments at 3% each.
- Discussion and Q and A 8%: classroom performance and cold calls, simulating real business challenges.
- Midterm exam 16%: programming project mid-term evaluation, business quantitative model evaluation.
- Final assessment 60%: programming project final evaluation, business quantitative model evaluation.

Textbooks

Main textbooks:

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (free online version available).
- Hands-On Large Language Models (free online version available).
- Quantecon.org (founded by Thomas Sargent) (free online).

Supplementary Textbooks

- Greene, Econometric Analysis, 8th Edition
- The Elements of Statistical Learning (Trevor Hastie, Robert Tibshirani, Jerome Friedman)
- Deep Learning, Goodfellow

Course Outline

- Week 1: What is an economics machine-learning model; overview of AI/ML models in finance; industry overview.
- Weeks 2–3: Machine-learning modeling workflow; how economic knowledge explains quantitative models; Python installation and fundamentals.
- Week 3: Practical financial data analysis; downloading official data via APIs and web scraping; large-scale financial data analysis (also the start of the final quantitative project).
- Weeks 5–6: Linear machine-learning model training: OLS, Lasso, Ridge, Elastic Net.
- Week 7 (Midterm): Design a linear ML model with real financial data.

Course Outline (cont.)

- Week 8: Classification algorithms (Quant risk and classification models)
- Weeks 9–10: Ensemble tree algorithms (XGBoost, Random Forest)
- Weeks 11–12: Neural network models (ANN + CNN)
- Weeks 13–15: Large models (Transformer)
- Weeks 16–17 (Final): Quantitative modeling in finance, model showcase, improvement, discussion, and critique.

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

Why Should we Economics Undergraduates Learn AI Models?

- Motivation for learning quantitative methods
- How these skills help future careers
- How to start learning

Machine-Learning Models are Dominating Various Areas

Machine-learning models provide algorithmic support and are the core of AI. Their powerful predictive ability has expanded into everyday life: text analysis (ChatGPT, BERT, Llama, DeepSeek), image recognition (facial recognition), speech recognition, recommendation systems (Taobao, JD, Amazon), medical research (accelerating COVID-19 vaccine development), and many more. In ****Finance and Economics****, many analyses and predictions have also been overtaken by ML, surpassing human analysts in both accuracy and speed.

Why ML Became a Darling of Financial & Economic Markets

So why now?

- Highly competitive market demands
- ML models can evaluate both value and risk
- Increasing market risk requires accurate investment and risk analysis (e.g., real-estate investments)

AI vs Machine Learning vs Deep Learning

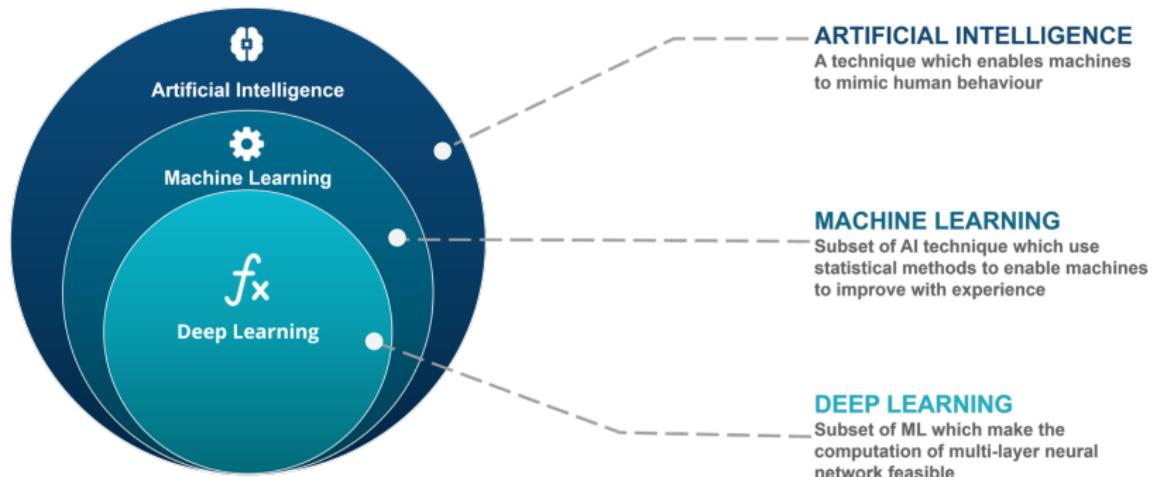


图: AI vs Machine Learning vs Deep Learning

source: <https://www.edureka.co/>

Common ML Algorithms in Economics

Economics typically uses:

- 1. Linear algorithms:** OLS, 2SLS, Logit, etc. These are foundational – BLUE (Best Linear Unbiased Estimation). For economics students, mastering econometrics and statistics is essential. Other linear methods: LASSO, Ridge, Elastic Net, MARS.
- 2. Ensemble learning:** XGBoost, CatBoost, AdaBoost, Random Forest.
- 3. Deep learning:** ANN (Artificial Neural Network), CNN, RNN.
- 4. Others:** Bayesian estimation, KNN, K-means, PCA, Causal Forest (Susan Athey), Double Machine Learning (Victor Chernozhukov), etc.

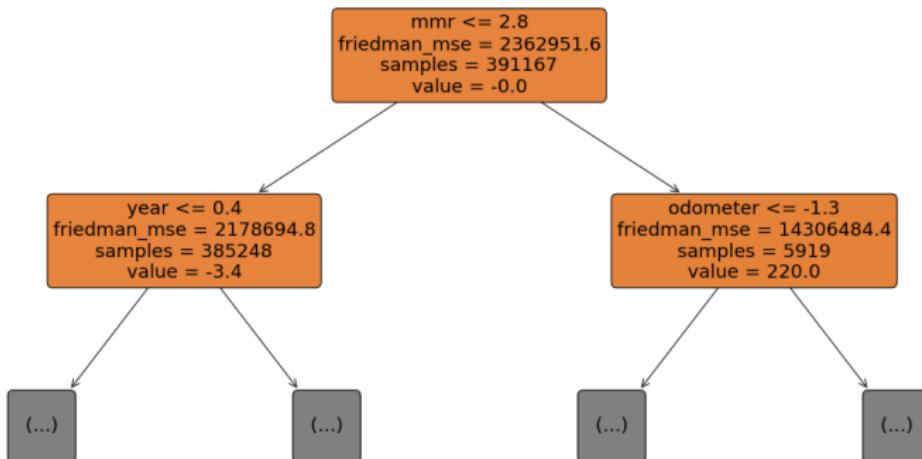
Machine Learning is not a Black Box

- In Python, building ML models is straightforward.
- Ensemble methods like XGBoost, CatBoost, AdaBoost simplify ensemble learning.
- TensorFlow and PyTorch streamline deep-learning model construction.
- HuggingFace Transformers, Unslot, LlamaFactory simplify large-model building.
- We'll show two simplified ML models (minimal architectures and parameters for learning purposes).

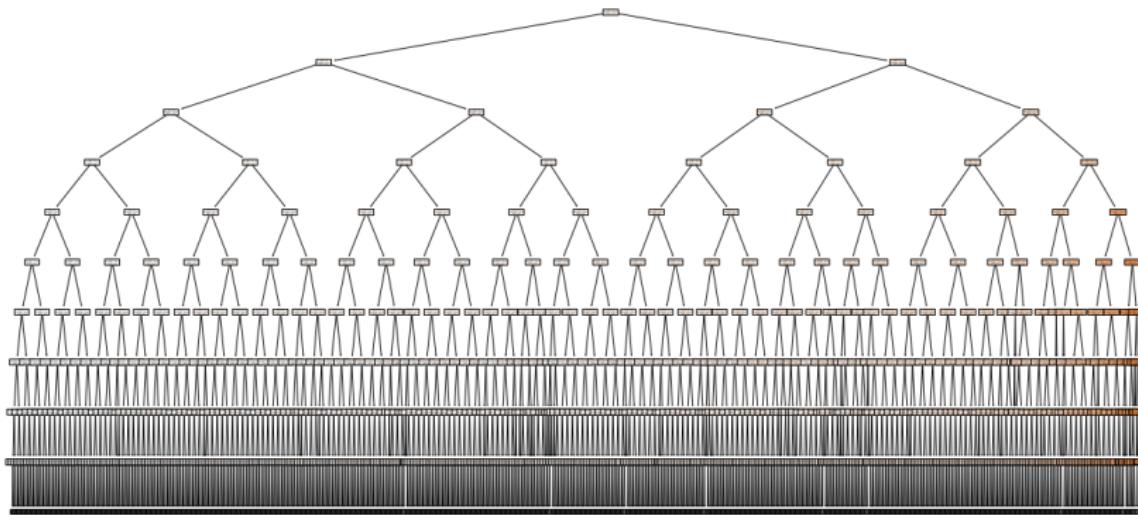
Ensemble Learning (Gradient Boosting Model Construction & Prediction)

```
hyperparameters_gb={'n_estimators':200,'learning_rate':0.1,  
reg_gb=GradientBoostingRegressor(**hyperparameters_gb)  
reg_gb.fit(xtrain_ann,ytrain_ann)  
ypred_gb = reg_gb.predict(xtest_ann)
```

Ensemble Learning (Gradient Boosting Framework)



Ensemble Learning (Gradient Boosting Framework)



Deep Learning (ANN Model Construction)

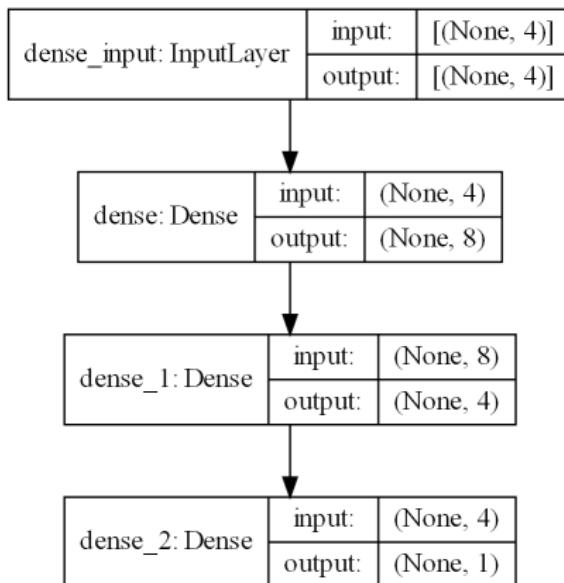
```
K.clear_session()
epoches=10
batch_size=128
model_ann = Sequential()
optimizer = keras.optimizers.Adam(lr=0.001)
model_ann.add(Dense(8,activation = 'relu',
                    input_dim = len(xtrain_ann.columns)))
model_ann.add(Dense(4,activation = 'relu'))
model_ann.add(Dense(1,activation ='linear'))
model_ann.compile(optimizer = optimizer,loss = 'mse',metrics = ['mae'])
model_ann.summary()
```

Deep Learning (ANN Training & Prediction)

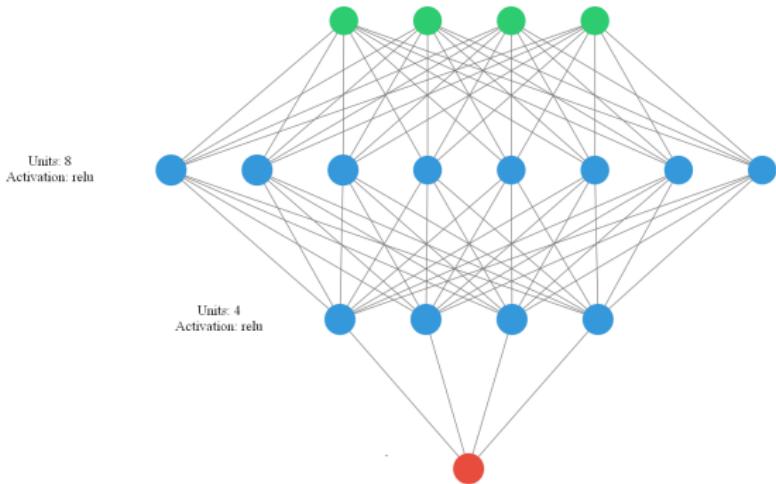
```
%%time
with tf.device('/gpu:0'):
    model_ann.fit(xtrain_ann,ytrain_ann,
                   batch_size=batch_size,epochs=epochs,
                   verbose=0)
ytest_ann = model_ann.predict(xtest_ann,batch_size = 32)
```

Wall time: 1min 29s

Deep Learning (ANN Framework)



Deep Learning (ANN Framework)



Artificial Intelligence GPT Model

```
from transformers import AutoTokenizer, AutoModelForCausalLM

# Load the tokenizer and model
tokenizer = AutoTokenizer.from_pretrained("gpt2")
gpt_model = AutoModelForCausalLM.from_pretrained("gpt2")
# Example input text
input_text = "what is python"
# input_text = "Python example of using bert for word embedding"

# Tokenize the input text
inputs = tokenizer(input_text, return_tensors="pt")

# Generate text
outputs = gpt_model.generate(**inputs, max_new_tokens=50)

# Decode the generated text
generated_text = tokenizer.decode(outputs[0], skip_special_tokens=True)
print(generated_text)
```

Setting `pad_token_id` to `eos_token_id`:None for open-end generation.
what is python?

Python is a programming language that is used to create and manipulate data.

Short History of Machine Learning

1. Linear models: Legendre (1805) and Gauss (1809) used linear regression to predict planetary motion.
2. Early deep-learning: Warren McCulloch & Walter Pitts (1943) created a computational neural-network model.

source: Timeline of AI (Wikipedia)

Cases where ML Surpassed Humans

1. AlphaGo (Microsoft): beat world champion Lee Sedol in 2016.
2. GitHub Copilot (Microsoft): AI writes code on demand, completing junior-coder tasks.
<https://copilot.github.com/>
3. ChatGPT (Microsoft): automatically generates articles, code, replacing finance analysts, economists, programmers.
4. Future: $AI \rightarrow AI \rightarrow AI \rightarrow \dots$

What Cannot be Replaced by AI?

- Independent thinking
- Creativity
- In the AI era, the core competition is for young talent with independent thinking and creativity (just like you all).

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

What is Learning? The Human Societal Learning Process

We can compare machine learning with human learning.

1. Learning is not a new concept.
2. Human evolution itself is a learning process.
3. Learning = Try + Improve + Repetition.

Human Learning is Slow and Accumulative

The only lesson history teaches us is that we do not learn from history. (“We learn from history that we do not learn from history.”) — Hegel



图: Georg Wilhelm Friedrich Hegel

Human Learning Issues

- From the ML perspective, Hegel's point means society's learning rate is too low, i.e., η is too small.
- Note: learning rate is a key hyperparameter in ML, controlling how fast the algorithm learns.

ML Advantages Over Human Learning

1. ML mimics human learning via algorithmic iteration: try + improve + repeat.
2. The difference is the speed of repetition and optimization: a human may spend months or years per cycle; a machine can do it in less than a second.
3. Experienced financial analysts rely on decades of work and thousands of cases, yet ML can analyze hundreds of thousands—or even millions—of assets in a few hours, constantly improving its predictive power. Thus ML far surpasses human analysts in both accuracy and efficiency.

How ML Learns from Data

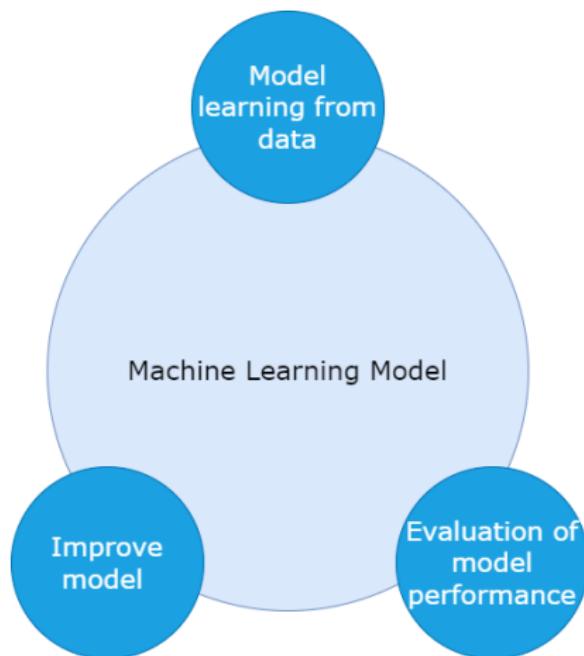


图: How ML Learns from Data

ML Models in Economics

1. ML models in finance/ economics—quantitative modeling (Quant)—are widely used in stock investment, high-frequency trading, asset management, risk management, asset valuation, insurance, fraud detection, etc.
2. Many of our economics peers aspire to become quant researchers, data scientists, etc.
3. For economists, mastering Python-based quantitative modeling is essential.

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

Why Quant Models Are Basic Skills for Economists?

1. Economists' simple analyses are no longer sufficient (e.g., clients want to know not just the price rise but why, and by how much).
2. Quant models run far faster than human analysis (e.g., real-estate price analysis, stock fundamentals, price forecasting, risk forecasting, report generation).
3. High-quality quant models (including ML and deep-learning variants) often match or exceed human analytic accuracy.

⇒ Quant models are the future essential skill for economists.

Job Market Demand in Economics and Finance

The best way to gauge market trends is to see what talent is needed. On Glassdoor we searched quantitative modeling demands and requirements.



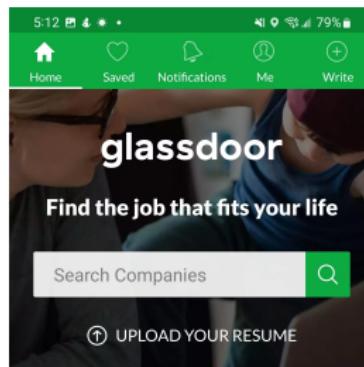
Get notified of new jobs

Tired of searching for jobs? Create a job alert to see the freshest jobs daily

[Create Job Alert](#)

Suggested Searches

quant
Townshipin Cities Amf, MN (US)



Get notified of new jobs

Tired of searching for jobs? Create a job alert to see the freshest jobs daily

[Create Job Alert](#)

Suggested Searches

quant
Townshipin Cities Amf, MN (US)

Quant Talent Demand Cases 1, 2

1:15 4 * 73% ← Quant Research and Model ... Morgan Stanley • New York, NY ⚡

Job Company Rating Why V

QUALIFICATIONS

7-10 years of experience in the financial services industry in a quantitative field, preferably with experience in model development/review, risk modeling and portfolio optimatio

At least Master's Degree in a technical field such as Mathematics, Statistics, Econometrics or Operations Research

Existing experience in leading conversations in Firm Risk Committees as well as with Model risk management function is preferable

Programming skills in statistical packages such as R, python or Matlab and familiarity with database systems such as Sybase, MS SQL

Familiarity with vendor risk systems such as RiskMetrics, BlackRock Aladdin, MSCI/

1:50 4 * 71% ← Quant Analyst Barclays PLC • New York, NY ⚡

Job Company Rating

What we're looking for:

- Bachelor's Degree in quantitative discipline e.g. Finance, Mathematics/ Statistics or Economics
- Experience in statistical model development
- Knowledge of data analysis, theory and statistical techniques
- Proficiency with analytical software Python, R, SQL tools e.g., Oracle, Unix platforms, and Microsoft Office

Skills that will help you in the role:

- Graduate studies, Masters or PhD in quantitative discipline
- Analytical work experience in a financial services company and strong technical and / or statistical skills with proven ability to process large datasets into meaningful



Quant Talent Demand Cases 3, 4

OPTIONS DESK QUANT
XR Trading, LLC • Remote

| Job | Company | Rating |
|-----|---------|--------|
|-----|---------|--------|

Required Skills

- Strong mathematical aptitude
- Fluency in statistical methods and modeling
- Coding skills in Python (C++ familiarity is a plus)
- Exposure and familiarity with machine learning techniques
- Values teamwork but capable of thinking independently
- Works on own Initiative / hustle – takes a pragmatic approach
- Effective communication with all levels of professional experience
- Ability to retain information and then teach others what you have learned

Required Experience

Data and Applied Scientist
Microsoft Corporation • Redmond, WA

| Job | Company | Rating |
|-----|---------|--------|
|-----|---------|--------|

Required:

- A bachelors or higher degree in computer science, machine learning, statistics, math, economics, business or other scientific or quant-focused field

- Programming skills (esp. related to data technologies like Python, Java, C#, etc.)
- 2 or more years of experience using data/ML/AI to impact critical product or business decisions

Preferred:

- Experience with hypothesis testing, graph theory and experiment design
- A proven track record of collaborating across organizational boundaries and

Quant Talent Demand Cases 5, 6

1:19 4 * • 73%
 ← Quant Research and Model ...
 Morgan Stanley • New York, NY

Job Company Rating Why V

QUALIFICATIONS

7-10 years of experience in the financial services industry in a quantitative field, preferably with experience in model development/review, risk modeling and portfolio optimizatio

At least Master's Degree in a technical field such as Mathematics, Statistics, Econometrics or Operations Research

Existing experience in leading conversations in Firm Risk Committees as well as with Model risk management function is preferable

Programming skills in statistical packages such as R, python or Matlab and familiarity with database systems such as Sybase, MS SQL

Familiarity with vendor risk systems such as RiskMetrics, BlackRock Aladdin, MSCI/Barron Yield Book, Barclays POINT, SunGard

1:16 4 * • 73%
 ← Quant Analytics Associate
 KeyCorp • Cleveland, OH

Job Company Rating Why V

REQUIRED QUALIFICATIONS

- Bachelor's degree (or its equivalent) in statistics, mathematics, economics, financial engineering, data sciences, predictive modeling, or other quantitative disciplines and at least 1 year of relevant experience; 0 with Master's or PhD

DATA LITERACY

- Understanding of:
Data wrangling including information documentation and importing data from different formats

- Data wrangling including information documentation and importing data from different formats

- Descriptive statistics, random variables, common distributions, outliers

Key Keywords for Quant Modeling Jobs

From the above postings we see two tightly related keywords:
Economics and Python.

- Economics: In today's financial markets, economics is often classified as a STEM discipline, requiring strong quantitative, statistical, and programming skills. Economics students' math-stat abilities are expected to match those of math or statistics majors. Good programming ability is also essential.
- Python: The most popular language for quantitative modeling today.

Knowledge Base for Economic Quantitative Modeling

1. **Economics:** the social science concerned with production, distribution, and consumption of goods and services.
2. **Quantitative modeling:** using econometrics & ML to build models that explain and solve economic problems.
3. **Python:** the trending programming language for building quantitative models.

Economic AI Modeling: an Organic Combination of Three Knowledge Areas

This course content goes beyond Python alone:

- Having Python skills does ****not**** equal economic forecasting.
- Applying economic theory does ****not**** equal quantitative modeling.
- Economic AI modeling = Python + Economics + ML models.

Industry Quant Modelers: Specific Roles & Workflow

Building quant models is complex with stringent accuracy and quality controls. Model quality management is even more critical than the algorithm itself.

1. First-line modeler: gathers business requirements, obtains data, builds the quant model, and hands it to developers for application.
2. Second-line modeler (model validating): reviews first-line models, identifies issues, and suggests improvements.
3. Third-line modeler (model auditing): ensures compliance, legal, and regulatory adherence.

Audience for Learning Quantitative Modeling

Targeted at:

1. Those wishing to work in industry as quant researchers, data scientists, etc.
2. Those wishing to pursue academic research (PhD in Economics/Finance, MSc in Financial Engineering).
3. Those interested in policy making research.
4. Even those intending to start their own ventures.

Course Schedule

What is a Financial AI Model?

Advantages of ML Models

Demand in Economics and Finance Jobs

Learning & Practice Resources

Boosting Skills: Reading White Papers & Research Papers

- Model documentation helps you quickly master a model (e.g., <https://xgboost.readthedocs.io/en/stable/tutorials/model.html> or its offline PDF). Source: model website or GitHub.
- Papers delve deeper into internal structures (e.g., <https://arxiv.org/abs/1603.02754>). Find them via Google Scholar or arXiv.

If You Encounter Coding Issues

- Search engines such as Bing
- StackOverflow
- Python's help function
- ChatGPT, Deepseek, Copilot, Cursor

Example of a Coding Issue

give me python template of using xgboost to predict stock price, give me python codes only

引用 7 篇资料作为参考 >

python

 下载  复制  运行 | 

```
# -*- coding: utf-8 -*-
"""
XGBoost Stock Price Prediction Template

This template provides a basic structure for predicting stock prices using XGBoost.
Remember to install required packages: pip install xgboost pandas numpy yfinance scik:
"""

import pandas as pd
import numpy as np
```



Data Resource 1: Built-in scikit-learn Data

Advantages: convenience, one-line code. Disadvantages: small size, not suitable for realistic modeling, only for simple demos.

sklearn.datasets: Datasets

The `sklearn.datasets` module includes utilities to load datasets, including methods to load and fetch popular reference datasets. It also features some artificial data generators.

User guide: See the [Dataset loading utilities](#) section for further details.

Loaders

| | |
|---|---|
| <code>datasets.clear_data_home([data_home])</code> | Delete all the content of the data home cache. |
| <code>datasets.dump_svmlight_file(X, y, f, *, [, ...])</code> | Dump the dataset in svmlight / libsvm file format. |
| <code>datasets.fetch_20newsgroups(*[, data_home, ...])</code> | Load the filenames and data from the 20 newsgroups dataset (classification). |
| <code>datasets.fetch_20newsgroups_vectorized(*[, ...])</code> | Load and vectorize the 20 newsgroups dataset (classification). |
| <code>datasets.fetch_california_housing(*[, ...])</code> | Load the California housing dataset (regression). |
| <code>datasets.fetch_covtype(*[, data_home, ...])</code> | Load the covertype dataset (classification). |
| <code>datasets.fetch_kddcup99(*[, subset, ...])</code> | Load the kddcup99 dataset (classification). |
| <code>datasets.fetch_lfw_pairs(*[, subset, ...])</code> | Load the Labeled Faces in the Wild (LFW) pairs dataset (classification). |
| <code>datasets.fetch_lfw_people(*[, data_home, ...])</code> | Load the Labeled Faces in the Wild (LFW) people dataset (classification). |
| <code>datasets.fetch_olivetti_faces(*[, ...])</code> | Load the Olivetti faces data-set from AT&T (classification). |
| <code>datasets.fetch_opennl([name, version, ...])</code> | Fetch dataset from opennl by name or dataset id. |
| <code>datasets.fetch_rcv1(*[, data_home, subset, ...])</code> | Load the RCV1 multilabel dataset (classification). |
| <code>datasets.fetch_species_distributions(*[, ...])</code> | Loader for species distribution dataset from Phillips et. |
| <code>datasets.get_data_home([data_home])</code> | Return the path of the scikit-learn data directory. |
| <code>datasets.load_boston(*[, return_X_y])</code> | DEPRECATED: <code>load_boston</code> is deprecated in 1.0 and will be removed in 1.2. |
| <code>datasets.load_breast_cancer(*[,</code> | Load and return the breast cancer wisconsin dataset (classification). |

Data Resource 2: Kaggle Datasets

- Kaggle, a subsidiary of Google LLC, is an online community of data scientists and ML practitioners.
- URL: <https://www.kaggle.com/datasets> (offline PDF available).
- Advantages: real large data, relatively new, many finance-related datasets.

The screenshot shows the Kaggle Datasets homepage. On the left, there's a sidebar with navigation links: Create, Home, Competitions, Datasets (which is selected), Code, Discussions, Courses, More, and Your Work. The main content area has a search bar at the top. Below it, a section titled "Datasets" is described as a place to "Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating." It features a "New Dataset" button and a "Your Work" button. A "Trending Datasets" section follows, with a search bar and categories: Computer Science, Education, Classification, Computer Vision, NLP, Data Visualization, and Pre-Trained Model. Below this, four dataset cards are displayed:

| New & Ancient Temples - India | maroc avito car dataset | Heart Attack Treatment Payments By Hospital | Mushrooms |
|---|--|---|--|
| Koustubh · Updated 2 hours ago Utility 8.4 · 14 kB 1 File (CSV) | Soufiane Bouzefzal · Updated 3 hours ago Utility 7.4 · 942 kB 1 File (CSV) | Satish Ganesan · Updated 7 hours ago Utility 10.0 · 1 MB 1 File (CSV) | Derek Kuru · Williams · Utility 6.9 · 79 kB 8791 Files (other) |

At the bottom, there's a "Popular Datasets" section and a footer with navigation icons.

Why We Use Jupyter Lab for Python

- Install Python via Jupyter:
<https://www.anaconda.com/products/individual> and
<https://jupyter.org/try>
- Jupyter is the most popular Python work environment, widely used in finance, tech, and government.
- Facilitates large team & project collaboration (first-line Quant team, second-line Quant team, EE team, Business team).
- Enables cloud computing (enterprise: AWS, MS Azure, Databricks; personal: Google Colab
<https://colab.research.google.com/>)

Jupyter

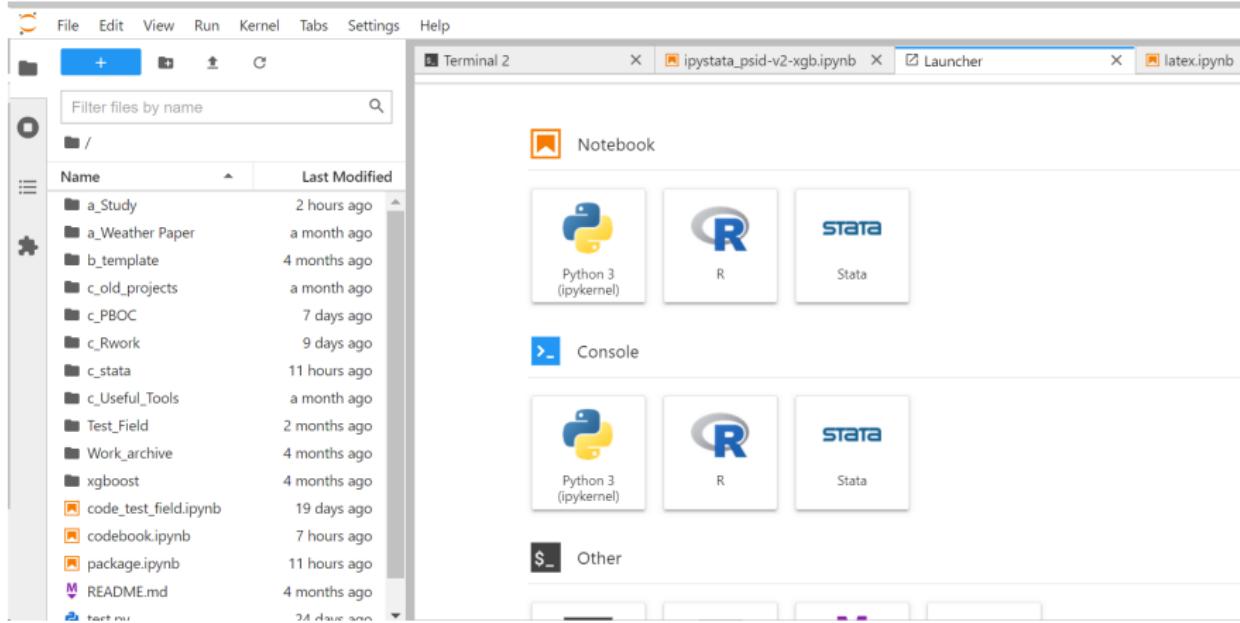


图: Jupyter Interface

Python Packages for ML Modeling

Install xgboost, scikit-learn, pandas, TensorFlow.

The screenshot shows the PyPI project page for xgboost 1.6.1. The top navigation bar includes links for Help, Sponsors, Log in, and Register. A search bar is at the top right. The main title is "xgboost 1.6.1". Below it is a button labeled "pip install xgboost". To the right, there is a green button labeled "Latest version" with a checkmark icon. The release date "Released: May 9, 2022" is also visible. The page content includes a "Project description" section with a "Project description" tab selected, showing the package version "xgboost 1.6.1". Other tabs include "Release history" and "Download files". The "Project links" section contains a "Homepage" link. The footer of the page includes standard navigation icons for back, forward, and search.

Making the Most of Your Learned Skills

Your decade-long training in English, econometrics, economics, programming, and especially creativity and wisdom should find an outlet.

- 1. Math skills**
- 2. English reading skills**
- 3. Coding skills**
- 4. Economics skills**
- 5. Young talent**

Related Literature

- Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.
- Leippold, M., Wang, Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/evaluation machine learning of treatment effects. *American Economic Review*, 107(5), 261–265.
- Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242.

Economic Quantitative Modeling Research Teams (Grouping)

- We'll assign groups.
- 3–4 people per group.
- Each group has a project lead.
- Most assignments are completed individually, including the final project.

Team Task: Question 1 — Install Anaconda, Python, and VS Code

- Troubleshoot collectively if issues arise.
- Differences: Windows vs macOS installation nuances; expect challenges.
- Optional: advanced users may try VS Code or WSL.
- Compile a brief report ($\leqslant 50$ words) on problems and solutions; to be reviewed and discussed in class next week.

Question 2 — Compute $1 + 1$

Each person should compute $1 + 1$ and save the .ipynb file.

The screenshot shows a Microsoft Edge browser window with a JupyterLab interface. The left sidebar displays a file tree with various notebooks and files. The main content area shows a code cell with the command `print(1+1)` and its output, which is the number 2. The browser's address bar shows the URL `localhost:8888/lab/tree/Untitled.ipynb`.

| Name | Last Modified |
|------------------------|---------------|
| a_Weather_Paper.ipynb | 4 months ago |
| c_old_projects | 10 months ago |
| c_PB0C | 7 months ago |
| c_Rwork | 9 months ago |
| c_stata | 3 months ago |
| c_Useful_Tools | 4 months ago |
| old | 4 months ago |
| Test_Field | 4 months ago |
| xgboost | 10 months ago |
| check_packages.ipynb | 3 months ago |
| codebook.doc | 5 months ago |
| codebook.ipynb | 3 months ago |
| CookBook_Metrics.ipynb | 4 months ago |
| package.ipynb | 4 months ago |
| README.md | 10 months ago |
| test.ipynb | 3 months ago |

Question 3 — Test NumPy Random Seed

```
import numpy as np  
  
np.random.seed(123)  
np.random.normal()
```

Question 4 — What do pandas, NumPy, scikit-learn, and matplotlib do?

- Pre-read: https://python-programming.quantecon.org/about_py.html
- Understand the purpose of each library. Cold calls will follow next week.

Question 5 — Register a GitHub Account

- Learn Git & GitHub, create an account at
<https://github.com/>
- We'll use Git for assignments, exams, and group collaboration.