

# 金融人工智能：未来经济学家的利器

Why AI & Machine Learning for Economists

葛雷

中国人民大学-数量经济

2025 年 7 月 1 日

网址：RUC AI沙龙

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

## ① 金融人工智能的时代

## ② 未来就业前景

### ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# 机器学习模型占领各个领域

机器学习模型（Machine Learning）是人工智能领域的算法支持与核心。因为机器学习模型的强大预测能力与准确程度，近些年来被用于生活中各种领域。比如：

- ① 文本大模型（ChatGPT, Bert, Llama, GLM）
  - ② 图像识别（人脸识别）
  - ③ 推荐算法（淘宝、京东、Amazon）
  - ④ 医药与生物（大大加快药物开发进程）
  - ⑤ 我们的 **经济学与金融学** 中的很多分析与预测，也在近几年逐渐被机器学习算法所占领

# 什么是金融人工智能建模

- 前沿：机器学习算法在金融领域的应用
- 交叉学科：数学+经济学+电脑编程
- 年轻人的舞台

# 为什么近些年机器学习成为金融经济市场的宠儿？

So, why now machine learning in economic & financial market?

- 高度竞争的经济金融市场需求
- 机器学习模型不但可以准确分析价值，还可以分析风险
- 市场风险加大，各大企业需要对投资与风险进行准确分析  
(想一想这些年的房地产公司)

# 机器学习算法占领金融经济市场

- 房地产
- 银行贷款风险
- 股票投资
- 资产管理
- 保险业
- 各个经济与金融行业

# 金融人工智能(Fintech)到底有多厉害

- 比人精度高
- 比人速度快
- 比人公平公正

# Example: 人工智能精度与速度都超过房产分析师

- 涉房不良资产包预测的误差仅**3%**（分析师误差10%）
- 单个房产价值预测的精准度约为5%（分析师误差8%）
- 每小时评估几万处房产（分析师1户/小时）

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# Quant行业与就业前景

- ① Quant领域：房地产、涉房资产风险、银行贷款审批、银行贷款风险管理、股市投资、资产管理、风险管理、资产估值、保险、反欺诈等各个金融行业
- ② Quant岗位：Quant Analyst, Quant Researcher, Quant Economist, Quant Modeler, Data Scientist and etc.

# Glassdoor是了解未来职业的一个很好的窗口

1:20 4 \* • 72% 

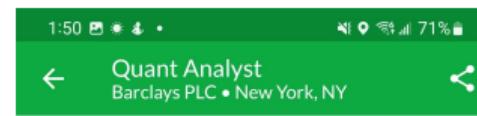
← Data and Applied Scientist  
Microsoft Corporation • Redmond, WA 

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

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:

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 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



# 业界量化模型师的具体分类与工作流程

量化模型的构建并不简单，有着严格的模型精准度与模型质量  
管理流程。模型的质量管理流程比算法本身更重要。

- ① 一线建模型研究员 (1st line modeler)：沟通Business获取需求，沟通数据组获得数据资源，通过数据与Business需求量化建模，沟通developer，将模型应用于business
- ② 二线建模型研究员 (Model validating)：负责检查一线模型的问题，提出模型的不足与提升意见
- ③ 三线建模型研究员 (Model auditing)：负责检查一线模型与二线模型的问题，检查模型合规与法律问题

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# AI vs Machine Learning vs Deep Learning

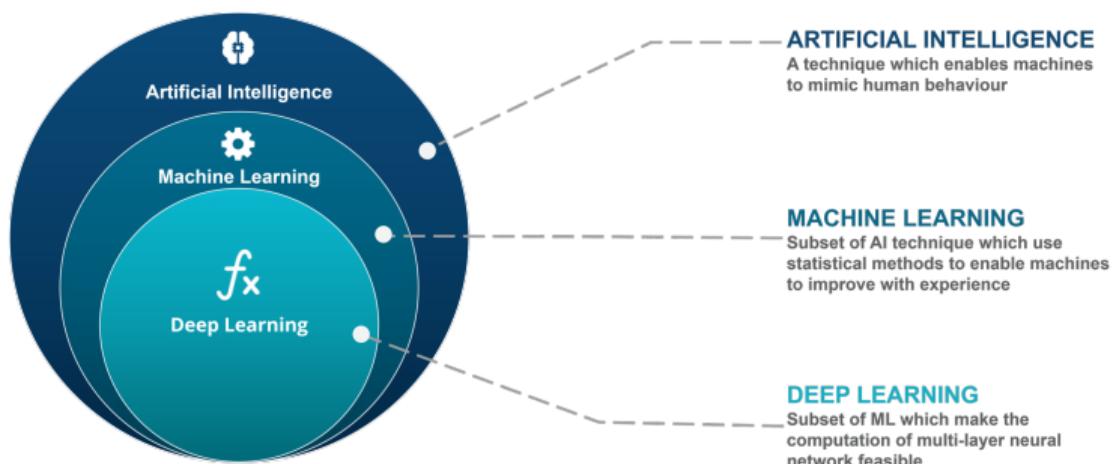


图 1: AI vs Machine Learning vs Deep Learning

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

# 机器学习在经济学上的常用算法

- ① **线性算法：** OLS, 2SLS, Logit等。OLS的兄弟们：LASSO, Ridge, Elastic, MARS (Multivariate adaptive regression spline)
- ② **集成学习（Ensemble Learning）：**  
Xgboost,Catboost,Adaboost,Random Forest
- ③ **深度学习（Deep Learning）：** Artificial Neural Network(ANN), CNN (Convolutional Neural Network), RNN (Recurrent Neural Netwrok), Transformer
- ④ **其他算法：** Bayesian Estimation, PCA, Autoencoder, Causal Forest (Susan Athey), Double Machine Learning (Victor Chernozhukov) 等等

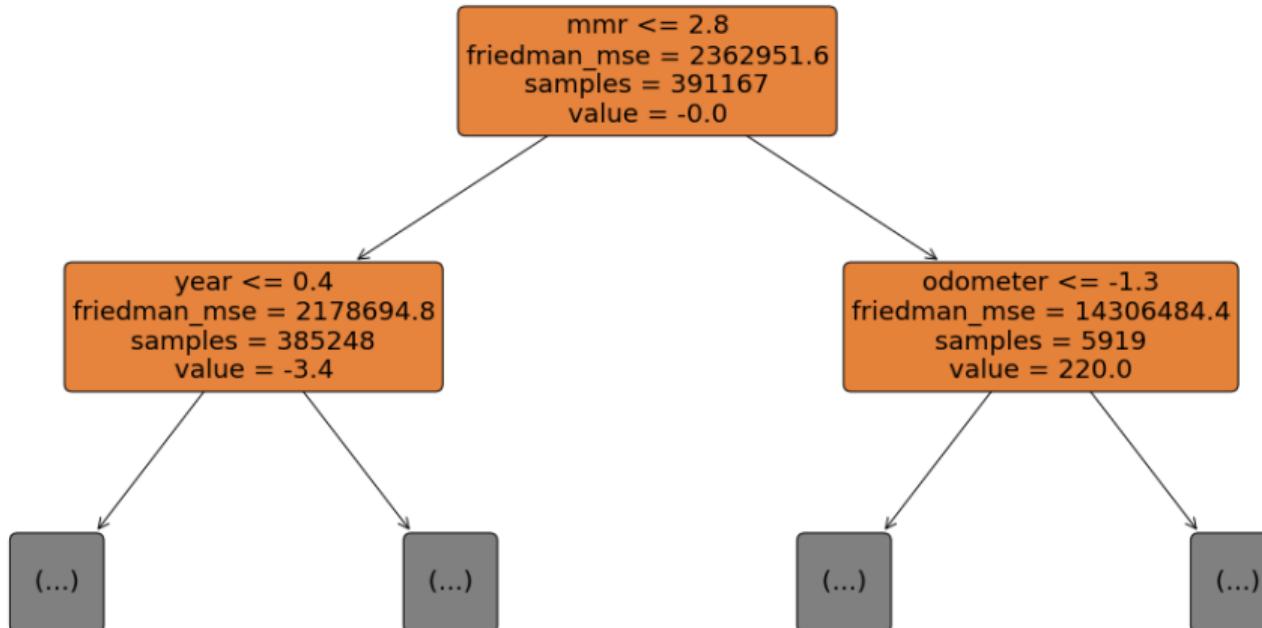
# Machine learning is not a black box

- 在python环境下，构建机器学习模型非常简单与直白
- 比如，TensorFlow将深度学习模型的构建简单化，Keras有进一步将TensorFlow简单化
- Xgboost, Catboost, Lightboost将集成学习简单化
- 下面我们展示两个简化的机器学习模型（为了学习需要模型结构和参数都为最简单设定）

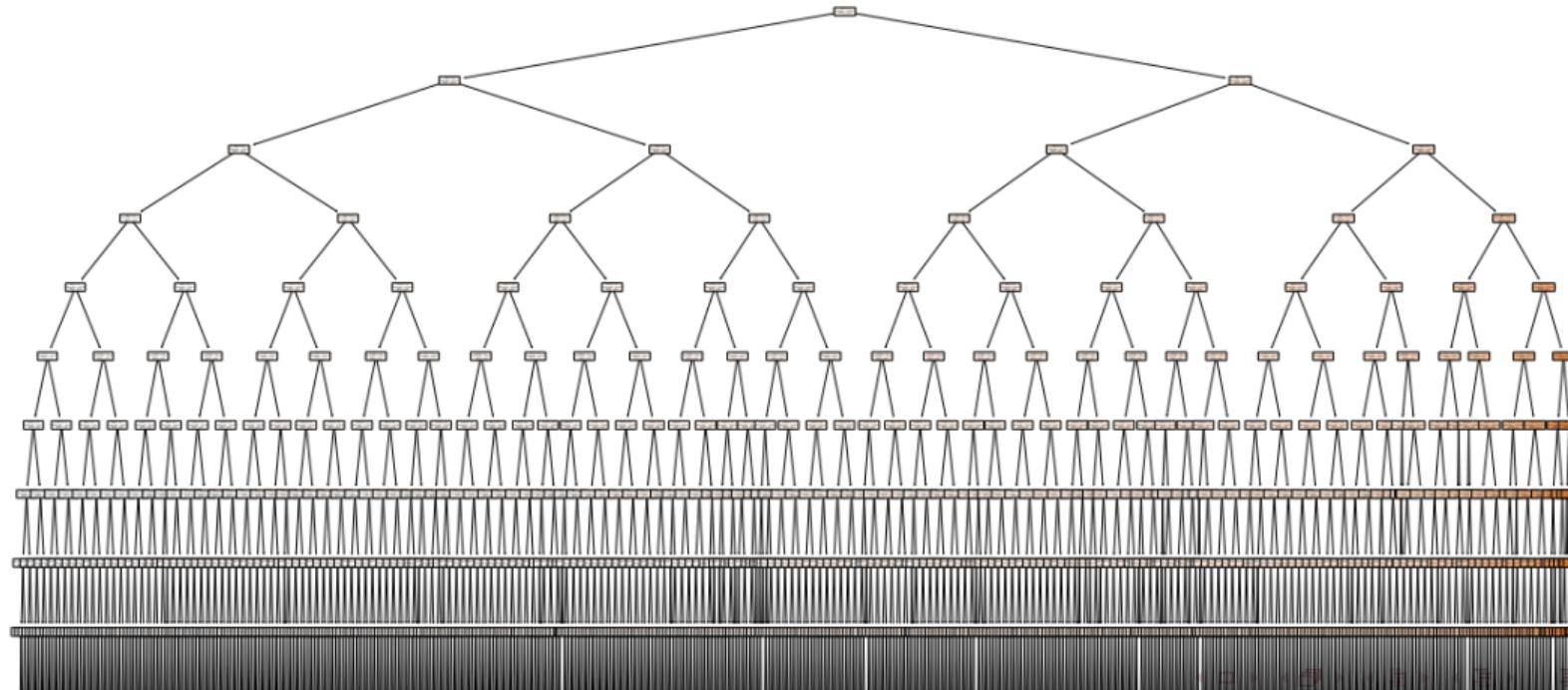
# 集成学习(Gradient Boosting 模型构建+预测)

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

# 集成学习(Gradient Boosting 框架)



# 集成学习(Gradient Boosting 框架)



# 深度学习（ANN 模型构建）

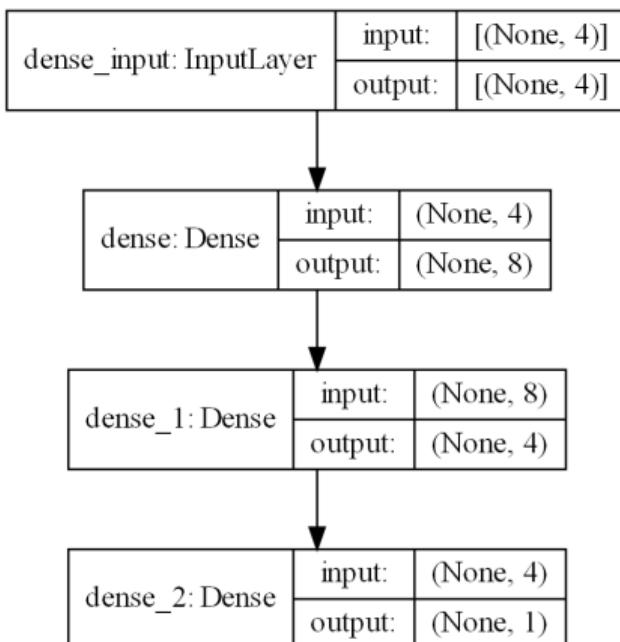
```
K.clear_session()
epochs=10
batch_size=128
model_ann = Sequential()
optimizer = keras.optimizers.Adam(lr=0.001)
model_ann.add(Dense(8,activation = 'relu',
                    input_dim = len(xtrain_ann.columns)))
model_ann.add(Dense(4,activation = 'relu'))
model_ann.add(Dense(1,activation ='linear'))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

# 深度学习（ANN 模型训练与预测）

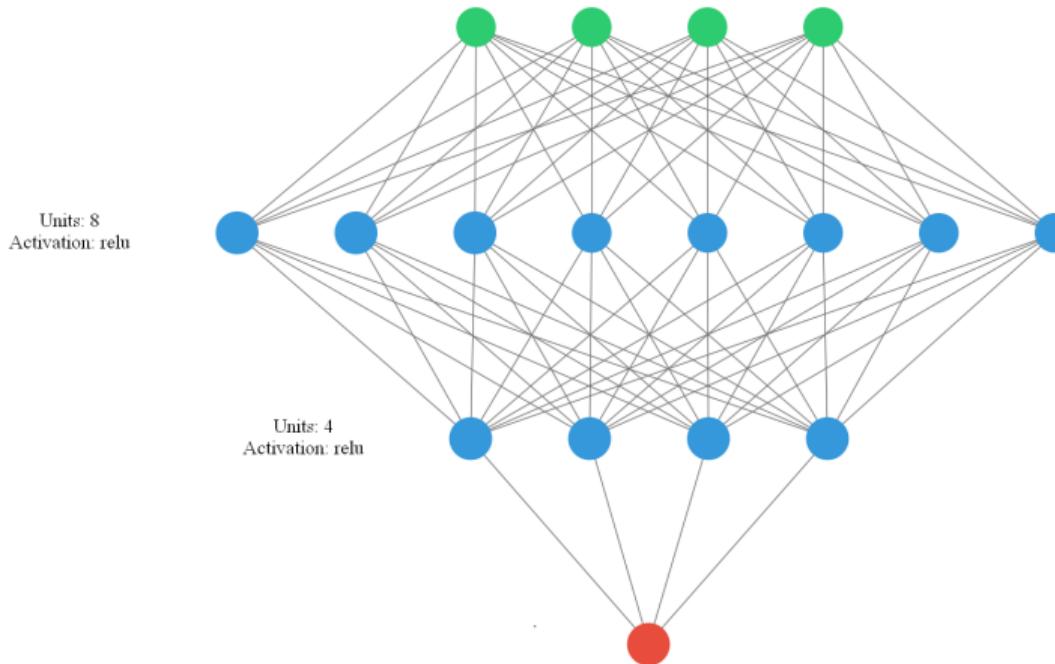
```
%%time
with tf.device('/gpu:0'):
    model_ANN.fit(xtrain_ANN,ytrain_ANN,
                   batch_size=batch_size,epochs=epochs,
                   verbose=0)
ytest_ANN = model_ANN.predict(xtest_ANN,batch_size = 32)
```

Wall time: 1min 29s

# 深度学习（ANN 模型框架）



# 深度学习（ANN 模型框架）



# Short history of machine learning

- ① 线性模型的创立: Legendre (1805) and Gauss (1809) used the linear regression for the prediction of planetary movement.
- ② 深度学习的雏形: Warren McCulloch and Walter Pitts (1943) opened the subject by creating a computational model for neural networks.

source: History of the AI from Wikipedia

# 机器学习算法战胜人类的案例

- ① AlphaGo: AlphaGo research project was formed around 2014 and beat world champion Lee Sedol 9th Dan in 2016.
- ② Github Copilot: 人工智能算法按照要求自动写代码,完成初级程序员的工作。 <https://copilot.github.com/>(点击链接注册可用)
- ③ AlphaCode: 人工智能算法按照要求自动生成算法，取代高级程序员的工作。 AlphaCode achieved an estimated rank within the top 54% of participants in CodeForces.  
<https://alphacode.deepmind.com>

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# 什么是学习：人类社会的学习过程（Human Society in Learning）

我们可以对比机器学习与人类的学习：

- ① 学习并不是一个新的概念
- ② 人类与人类社会的演进也是一个学习的过程
- ③ 人类学习=尝试+改进+不断重复

# 人类的学习是一个缓慢积累的过程 (long and slow process)

人类从历史学到的唯一的教训，就是人类没有从历史中吸取任何教训。（We learn from history that we do not learn from history.） ——黑格尔



图 2: Georg Wilhelm Friedrich Hegel

# 人类的学习是一个缓慢积累的过程 (long and slow process)

人类历史上漫长与痛苦的学习过程：

- ① 商有九世之乱：用了整整九个王，学习到了父死子继比兄终弟及更利于封建王朝的稳定统治 (中丁、外壬、河亶甲、祖乙、祖辛、沃甲、祖丁、南庚、阳甲)
- ② 斯图尔特王朝：用了整整一百年的时间学习并从封建集权制过渡到资本主义议会制度 (House of Stuart, 1603年至1714年，詹姆斯一世，查理一世，护国主克伦威尔...)
- ③ 高考刷题：五年高考三年模拟，高中三年不断尝试与改进的过程

# 人类学习的问题

- 用机器学习算法的角度总结黑格尔的话，就是人类社会的学习效率太低太慢。也就是Learning rate ( $\eta$ ) is too low。
- 注： Learning Rate是机器学习模型中的一个重要的超参数 (Hyperparameters)，控制着机器学习算法的学习速度

# 机器学习相对人类学习的优势

- ① 机器学习用算法高度模拟人类学习的过程
- ② 和人类一样，机器学习算法的秘诀也是：尝试+改进+不断重复

# How machine learning from the data

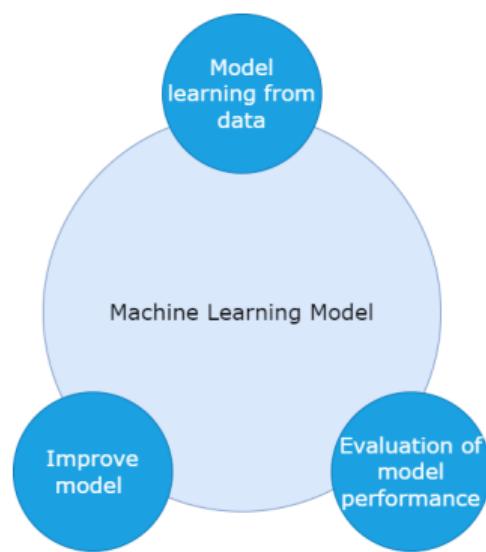


图 3: How machine learning from the data

# 机器学习相对人类学习的优势

- ① 人类资深金融分析师通过十几年的工作经验与几千个工作的案例不断重复总结经验
- ② 机器学习算法只需要几个小时就可以轻松分析几十万个甚至上百上千万个资产

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# 教材

- Quantecon.org (founded by Thomas Sargent)
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow(<https://github.com/ageron/handson-ml>) offline
- Greene, Econometric Analysis, 8th Edition

# 遇到问题：读模型说明书（White Paper）与论文（Paper）

- 模型的说明文件可以帮助同学们快速熟练使用模型（如：<https://xgboost.readthedocs.io/en/stable/tutorials/model.html offline>）。渠道：模型的官网 or github。
- 论文可以去更深层次的去了解模型的内部结构（如：<https://arxiv.org/abs/1603.02754>）。渠道：google scholar, Arxiv

# 遇到不会的问题怎么办 (Coding Questions)

- AI tools such VScode copilot, Cursor, Windsurf
- Search Engines such as Bing
- StackOverflow
- help function in Python

# 遇到不会的问题怎么办 (For Example)

Step1: Use key words to search your coding questions

stack overflow pandas groupby sum

全部 图片 新闻 视频 更多 工具

找到约 509,000 条结果 (用时 0.54 秒)

<https://stackoverflow.com/questions/10361407/how-do-i-pandas-group-by-to-get-sum> 翻译此页

**How do I Pandas group-by to get sum? - Stack Overflow**

2022年1月25日 — This operation will calculate the total number in one group with function `sum` , the result is a series with the same index as original `dataframe`. `df['Number'] = ...`

9 个回答 · 最佳答案: Use GroupBy.sum: `df.groupby(['Fruit','Name']).sum()` Out[31]: Number Fr...

Python pandas groupby aggregation - Stack Overflow 2014年9月29日

Pandas groupby.sum for all columns - Stack Overflow 2022年2月17日

Pandas : Groupby sum values - python - Stack Overflow 2021年8月17日

pandas groupby sum and agg [duplicate] - python - Stack ... 2020年6月16日

stackoverflow.com 中的其他相关信息

# 遇到不会的问题怎么办 (For Example)

Step2: find your answers or templates from the professional website such as the StackOverflow

9 Answers      Sorted by: Highest score (default) ▾

▲ Use [GroupBy.sum](#):

388 df.groupby(['Fruit','Name']).sum()

▼ Out[31]:

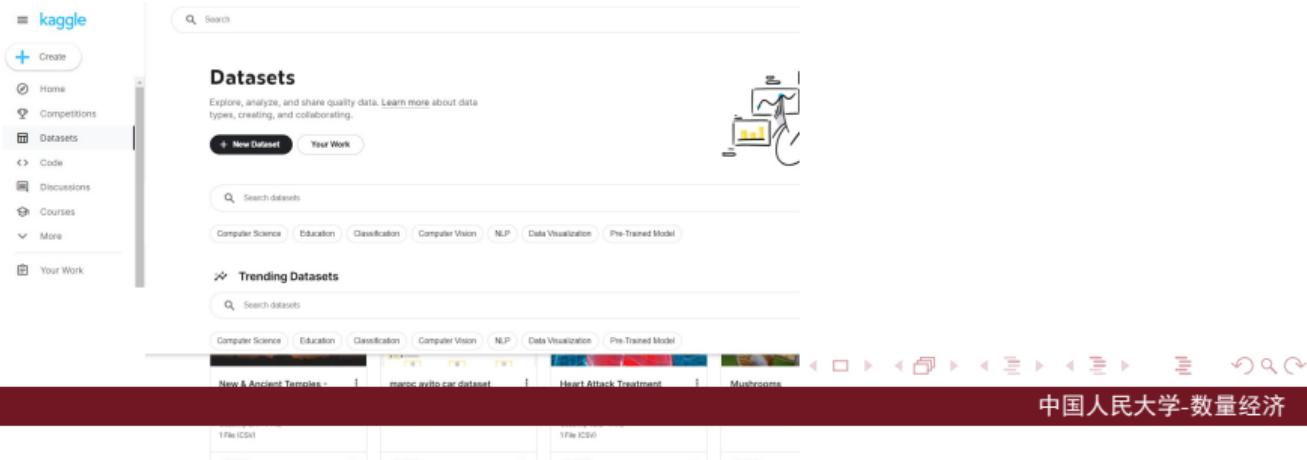
		Number
Apples	Bob	16
	Mike	9
	Steve	10
Grapes	Bob	35
	Tom	87
	Tony	15
Oranges	Bob	67
	Mike	57
	Tom	15

葛雷

中国农业大学-数量经济

# 数据与代码资源：Kaggle

- Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners.
- 网址：<https://www.kaggle.com/datasets> offline
- 优点：真实大数据、数据比较新、与经济与金融学相关数据比较多



# 提升量化建模的Secrets: 坚韧不屈的努力

## 人民大学第一届金融人工智能挑战赛

团队	分数	提交次数	最佳成绩提交时间	最后提交时间
张家玮的团队	89.473	131	2024/06/06 13:19	2024/06/06 13:19
莫湘源的团队	89.336	202	2024/05/22 01:11	2024/06/29 23:55
老虎机房产	88.627	84	2024/06/23 23:04	2024/06/23 23:04

# 不负平生所学

各位同学过去十几年勤学苦练的技能得到发挥。比如：英文能力、计量经济学能力、经济学能力、编程能力、更重要的是创造力与智慧。

- ① Math skills
- ② English reading skills
- ③ Economics skills
- ④ Coding skills
- ⑤ Young talents

## ① 金融人工智能的时代

## ② 未来就业前景

## ③ 什么是机器学习算法

## ④ 机器学习模型的优势

## ⑤ 学习建模方法

## ⑥ 附录：人大课程

# 人大的实战金融量化建模课程

- 人工智能与Python程序设计 (春、秋学期都有)
- 金融与经济实战量化建模
- 每周四下午2点到4点半，立德楼1002
- <https://datahub.ruc.edu.cn/>

# 人工智能与Python程序设计

中国人民大学 | 登录

< 人工智能与Python (金融量化24秋)

概况

课程

作业

资源库

人工智能与Python (金融量化24秋)

葛雷、吴亚聪 51

课程介绍

编辑

课程概述

机器学习模型 (Machine Learning) 是人工智能领域的算法支持与核心。因为机器学习模型的强大预测能力与准确程度，近些年来被各大金融机构所青睐，用来替代传统的经济分

葛雷

# 相关文献

- 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/neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261-65.
- 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.