

Week 4: 使用基础机器学习方法选取特征

在这一阶段，你需要使用基础的机器学习方法构建预测模型，并且通过训练出来的模型导出的feature importance list进一步筛选重要性极低的特征，最后对所有特征进行PCA降维，得到用于NN训练的特征组。

具体的步骤大概如下：

1. 使用 `TimeSeriesSplit` 对数据进行抽样，分离出train/test data
2. 使用 **Gradient Boosting Regression** 训练一个回归预测模型，RMSE作为loss
3. 从模型中获取feature importance列表，把importance极低的特征删除或者使用多项式等方法融合
4. 机器学习模型弃之不用
5. 假如特征维度过大，考虑特征进行适当的PCA降维，以提升模型训练速度
6. 将“原始数据输入→用于训练特征组输出”这一过程封装成函数，以供后续的实时预测使用

完成对NVDA的处理后，你需要额外获取以下股票代码的历史数据，以扩增数据量，获得更准确和泛化的模型：

```
['AAPL', 'INTC', 'AMD', 'TSM', 'QCOM', 'AVGO', 'SNPS', 'CRUS', 'META', 'AMZN',  
'MSFT', 'GOOGL', 'MU', 'MRVL', 'NXPI', 'STM', 'HPQ', 'IBM', 'CSCO']
```