Lifecycle of trading strategy development with machine learning

This is a 12-hour, in-depth, online workshop focusing on the challenges and nuances of working with financial data and applying machine learning to generate trading strategies. We will walk you through the complete lifecycle of trading strategies creation and improvement using machine learning, and ending with automated execution, with unique insights and commentaries from our own research and practice. We will make extensive use of Python packages such as Pandas, Scikit-learn, LightGBM, and execution platforms like QuantConnect. It will be co-taught by Dr. Ernest Chan and Dr. Roger Hunter, principals of QTS Capital Management, LLC. The course will be taught in 3 parts, each part will last 4 hours. Course schedule and registration can be found at www.epchan.com/workshops.

Course outline:

Part 1

A. Overview: Challenges of financial data science and machine learning
    a. Data cleansing: Why even simple daily data cannot be trusted.
    b. Features engineering: Claims that this step is easy for deep learning are *false*.
    c. Features selection: What even experts can get wrong here.
    d. Machine learning: shallow + deep learning work best together.
    e. Avoiding data snooping and selection bias: using CPCV.
    f. Metalabelling: improving your proprietary strategy without telling anyone.
    g. Backtesting: beyond machine learning.
    h. Automated execution: choosing a platform.

B. Data cleansing and features engineering
    a. Checking and adjusting price and volume data in stocks and futures.
    b. Survivorship bias and how to find it.
    c. Stationarity and "fractional differentiation".
    d. Sanity checks for news sentiment data.
    e. Sanity checks for earnings data.
    f. What is a security master and how to create one where none existed?
    g. Aggregating and encoding categorical data into features.

Part 2

C. Machine learning
    a. Better Start Simple: An example of simple features and shallow machine learning using logistic regression with L1 and L2 regularizations.

b. Deeper learning: Random forests and gradient boosted trees with Scikit-Learn and LightGBM.
c. Features selection using Mean Decrease Accuracy and SHAP: be careful where you apply that!
d. Cross validation and hyperparameters optimization.
e. Metrics for measuring machine learning outcomes.
f. Metalabelling: what common base models to use?

Part 3

D. *Backtesting*
a. Machine learning suggests, but does not determine, trading strategy.
b. Various ways of using the output of ML for trading.
c. Reduce data snooping bias: using simulations and CPCV.

E. *Automated Execution*
a. Using QuantConnect to automate strategies to trade on Interactive Brokers.