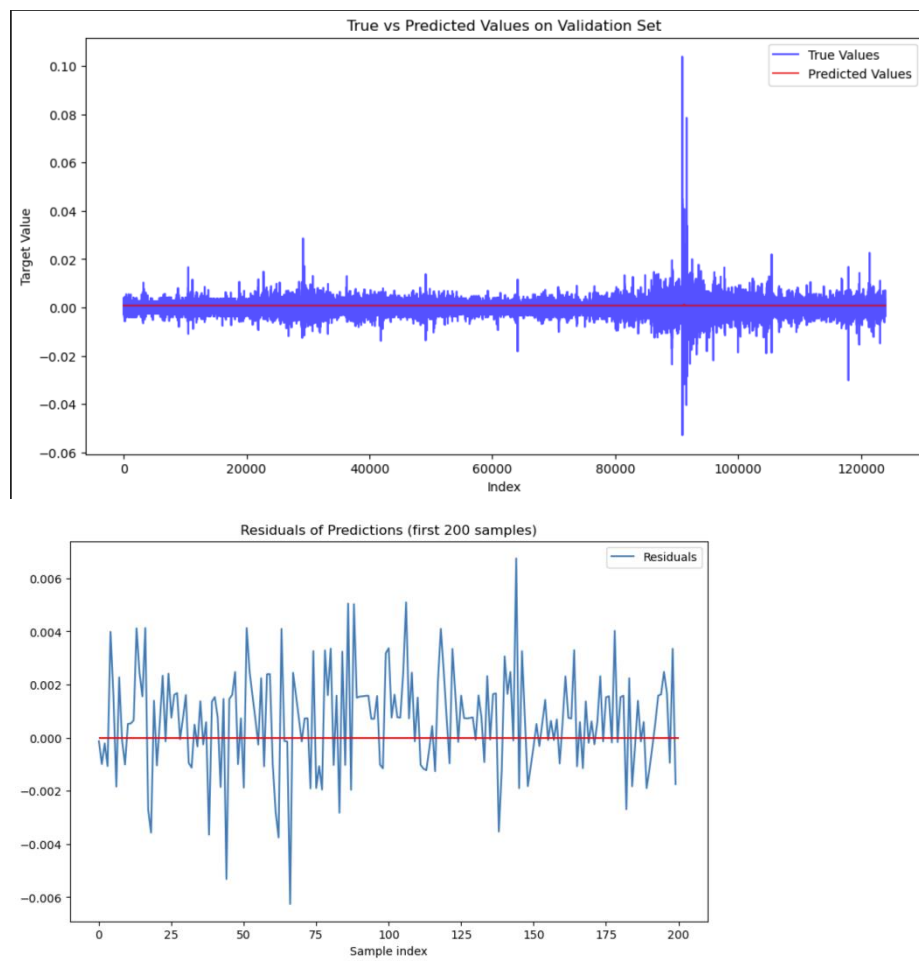
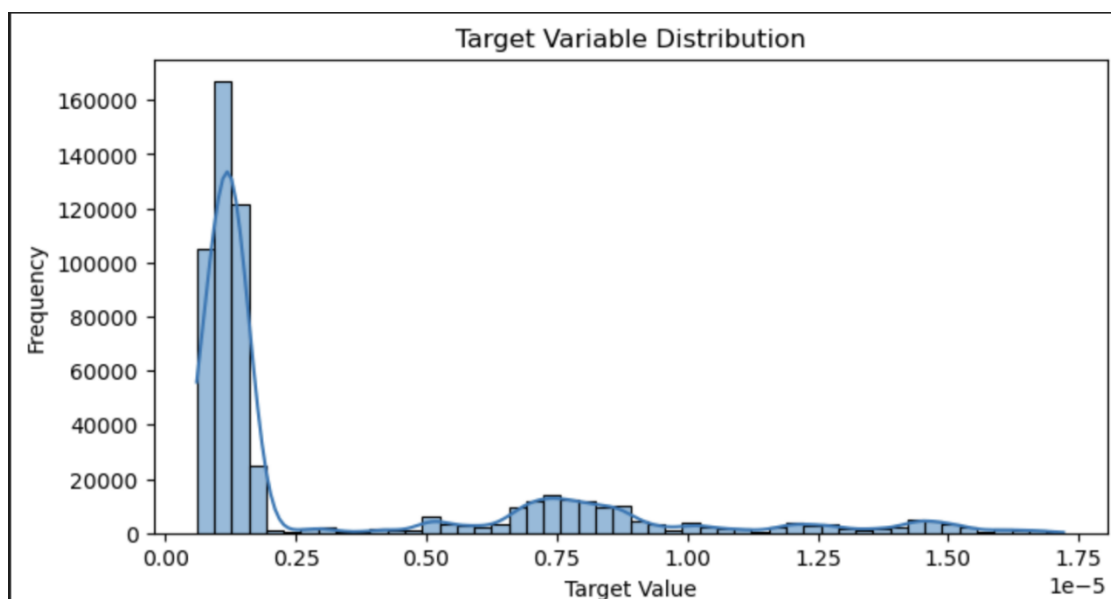


1. 原始数据集的 Return 和预测目标  $\text{Return.shift}(-1)$



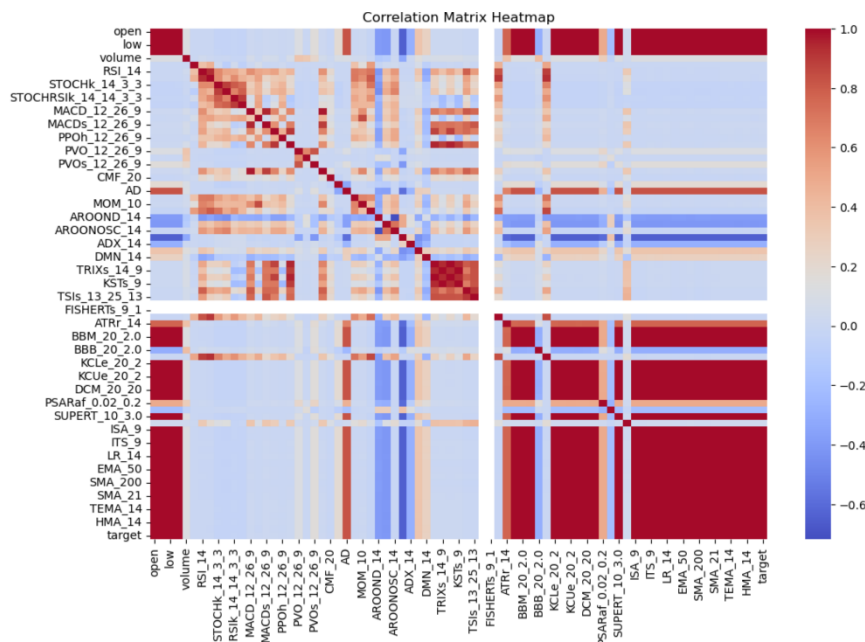
2. 指标使用: `ta_lib` 中可能有用的 72 个指标
3. 问题: Return 的波动性太大 (特别对于 1000PEPE 等类似的加密货币), 如果预测目标是非正态分布或在时间序列上不平稳 (特别是非常高频的数据), 那么机器学习和神经网络的预测方法可能不适用。
4. 分析: 使用 Close 作为预测目标, 可能会更加平稳



上图是 close.shift(-1)，也就是 target 的分布，发现它有偏态分布和厚尾两个特征。

5. 处理：

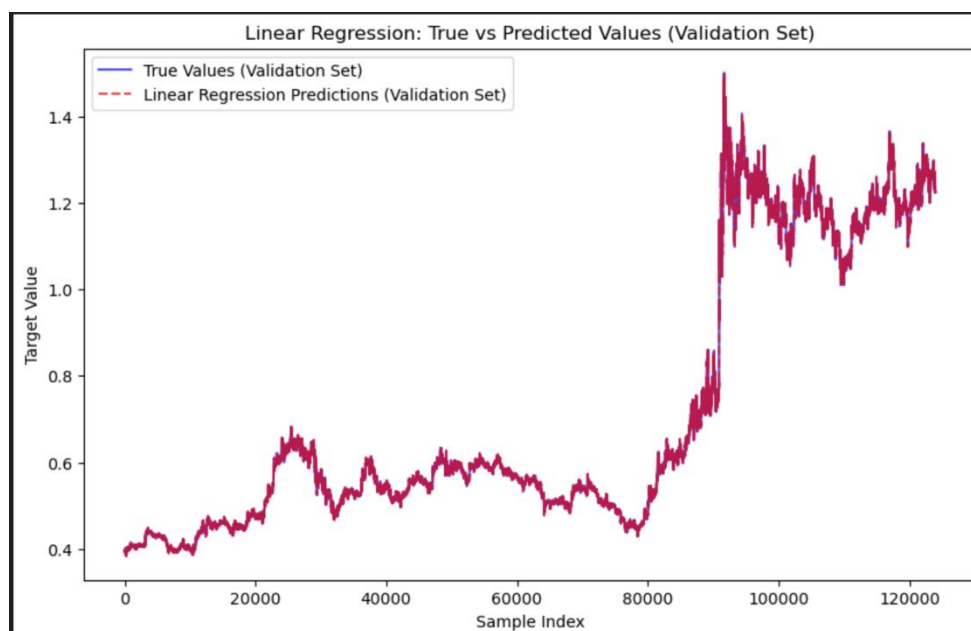
A . 由于指标较多（有 76 个），先用 correlation 计算其与 target 的相关性，绘制出热力图如下所示：



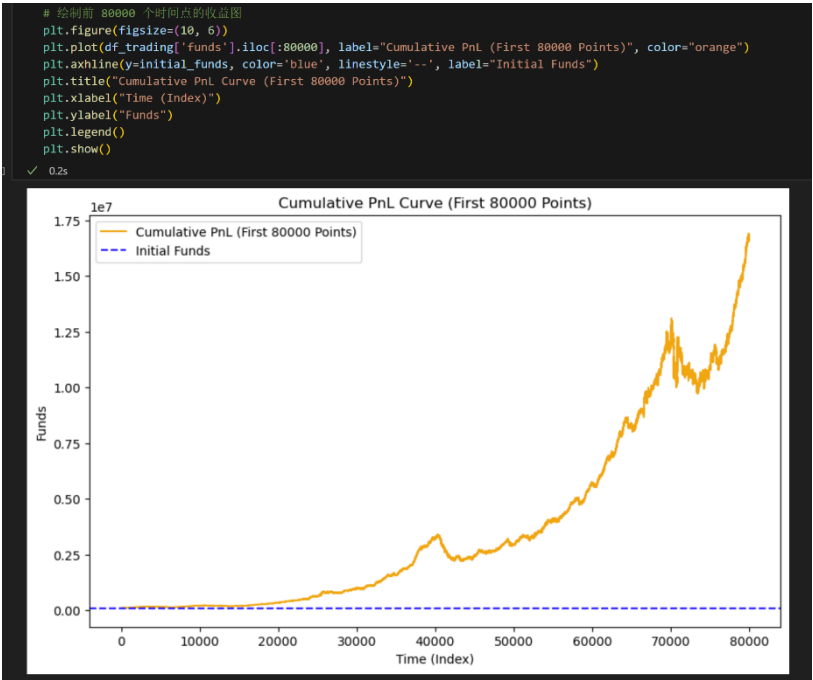
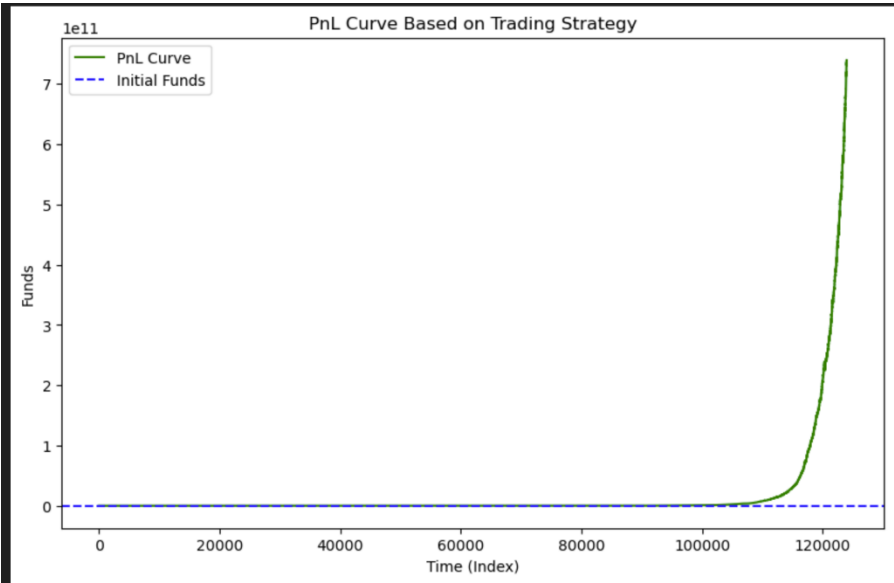
B . 根据热力图筛选出可能具有多重共线性的列然后删除，并根据阈值（绝对值>0.3）筛选出可能与 target 相关的特征列。

数据集大小：  
训练集：578421，测试集：123947，验证集：123948  
最终保留的特征列：['open', 'OBV', 'AD', 'AROOND\_14', 'AROONU\_14', 'CHOP\_14\_1\_100', 'ADX\_14', 'DMP\_14', 'DMN\_14', 'ATRr\_14', 'BBB\_20\_2.0', 'PSARaf\_0.02\_0.2']

C . 如果上述特征做得非常完备，可能不需要用神经网络，用简单的统计学模型即可预测。



D. 尝试运行的策略



PnL 如上所示，然后我之后观察一下是否使用未来数据、年化收益与 Sharpe、最大回撤以及其他的内容，再与您确认。