Home Work 2 due Sunday, Jan 29, before 11:59 pm

1. Insert a column in the data set where the entries are 1 if the stock outperforms SPY by more than 5% in the earnings period and -1 if it underperforms by -10% and 0 if the performance is between these values.

   Create both a DecisionTree classifier and a BaggingClassifier and use Optuna to find optimal hyperparameters.

   Use the period 2007-2009 to train both models on the data set constructed above and construct a set of features that optimizes the profit of the long-short strategy on the quarter 2010-04-01 – 2010-07-01

   Back test the performance of the models with the optimal features over the period 2010 – 2018.

   Compute Shapley values for the models and find the features with non-zero Shapley values. Train the models on the dataset with these features and back test as above

   Compute Sharpe Ratio, Information Ratio and alpha for the strategies and for the buy-and-hold strategy for SPY

2. We make a new definition of profit, instead of using an equal weighted portfolio, we weight the portfolio weights with the probabilities of the labels as follows: Use the bg_clf.predict_proba on the validation set, this will return an array of the probabilities $Prob_s(-1)$,$Prob_s(0)$, $Prob_s(+1)$ of the labels -1,0,+1 for each stock *s* so an array of dimension num_stocks x 3. We define the *'conviction'* of the model for each prediction as

   $$conviction_s = Prob_s(+1) - Prob_s(-1)$$

   and the $weight_s = conviction_s / \sum_{s'} \text{abs}(conviction_{s'})$ so $\sum_s \text{abs}(weight_s) = 1$

   The profit is then $\sum_s weight_s$ * 'next_period_return'

   Use Optuna and SHAP to find optimal hyperparameters and features, back test and compute Sharp Ratio etc