Machine Learning and Financial Trading

Rufus Ayeni

Overview

The idea of machine learning (ML) has been present for well over 60 years.

However, recently *(over the last decade)*, it has garnered a lot of attention *(from self-driving cars, to fraud detection, to product recommendations on online shopping platforms)*.

There is no doubt that ML has made considerable contributions in solving real-world problems.

Overview

Yet, in the area of trading and investing, there are mixed opinions about ML's usefulness.

Although institutions leverage ML to gain an advantage in the financial markets, many retail traders (*individuals who trade their own money via discount brokers*) have not experienced the same benefits for a couple reasons--

- Lack of knowledge (*some believe ML is difficult to understand, or that it simply offers no value*)
- Lack of resources (*institutions have millions of dollars to invest network infrastructure and to hire hundreds of PhDs to help gain an advantage in the markets*)

Problem Statement

Can machine learning enhance a retail trader's performance?

Data Wrangling

1. In our modeling, we used the price data of the Gold ETF (GLD).

2. The data was collected from Yahoo! Finance via its Python API.

3. The dataset consists of approximately 2700 observations from December 2011 to December 2021

4. Initial features were **`Date`**, **`Open`**, **`High`**, **`Low`**, **`Close`**, and **`Volume`**.

5. The data was remarkably clean (*no nulls, no duplicates, date was the correct data type,* `datetime`).

Data Wrangling - Feature Selection

1. Final feature set:

   `Close`: Closing price of GLD

   `RSI` (Relative Strength Index): Overbought or Oversold

   `ADX`: Non direction indicator- strength of trend (> 25)

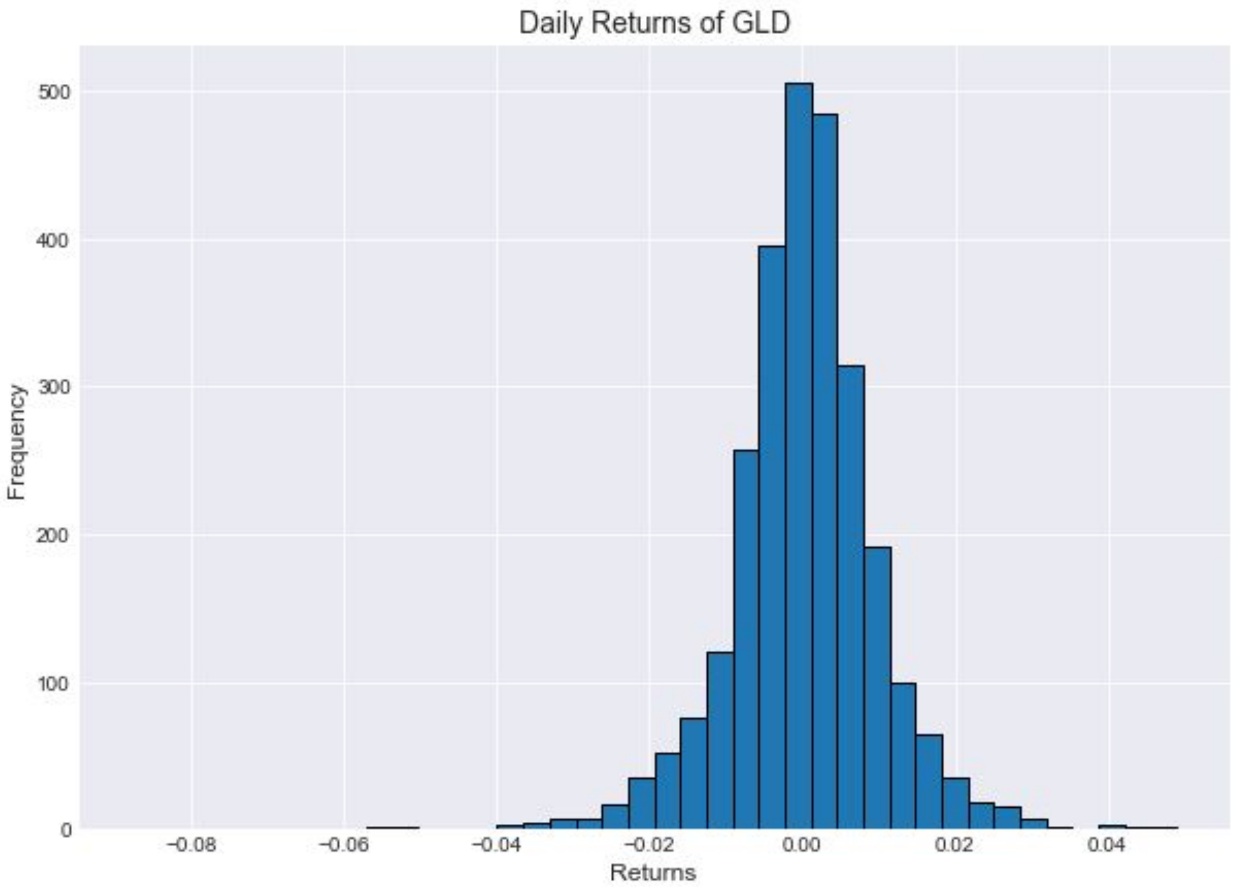
   `pSAR` (Parabolic Stop and Reverse): Trend following

   `TEMA` (Triple EMA): Trend following, without the lag

   `Daily_Return`: One day return of GLD

Exploratory Data Analysis (EDA) - Daily Closing Price of GLD ETF

EDA - Distribution of Daily Returns



count	2733.000000
mean	0.000133
std	0.009866
min	-0.087808
25%	-0.004792
50%	0.000407
75%	0.005108
max	0.049038

EDA - Mechanical System

In the EDA section, a simple BUY only strategy was created. The rules are:

Buy GLD ETF when the following are true:

```
RSI < 80 &
ADX > 25 &
Daily_Return.shift(1) > 0 &
Close > pSAR &
Close > TEMA
```

Exit GLD ETF when the following are true:

```
Close < pSAR &
Close < TEMA
```

The total returns of strategy: 6.31%

Modeling

The following models were used to generate buying signals:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier

Model Metrics

Model	Parameters	Accuracy		F1 score	
		Training	Testing	Training	Testing
Logistic Regression	no penalty max_iterations 1000	**0.525**	**0.510**	**0.586**	**0.576**
Random Forest Classifier	max_depth = 14	**0.639**	**0.522**	**0.689**	**0.537**
Decision Tree Classifier	max_depth = 5	0.574	0.557	0.648	0.614
Null	No BUY signal	0.479	0.479	UND	UND

Modeling Strategy Returns

Model	Model Strategy % Return	Percent Improvement over Mechanical System
Logistic Regression	35.21%	458%*
Random Forest Classifier	39.62%	527%*
Decision Tree Classifier	**42.73%**	**577%***

* Mechanical system percent return: 6.31%

Summary of Results

Initial tests show that ML can enhance a retail trader’s performance.

- Three models were created with accuracy rates between 51% and 55%.

- The three models improved strategy returns from 6% to 35%, 39%, and 42% respectively, representing and improvement of 458% to 577%.

Initially, the models’ performance metrics were poor when three years of data were used. After collecting 10 years of data, the metrics improved considerably.

Recommendations

To improve model metrics:

- use hyperparameter tuning for the aforementioned models
- explore additional models such as neural networks and principal component analysis

To improve returns:

- consider adding more features from the 100s of technical indicators available.
- change the observation time frame from daily to lower time frames such as hourly, 30m, 15m, etc to find more buying opportunities
- Test, test, test. This can't be emphasized enough. You must test extensively before going live *(i.e., placing live trades)*.