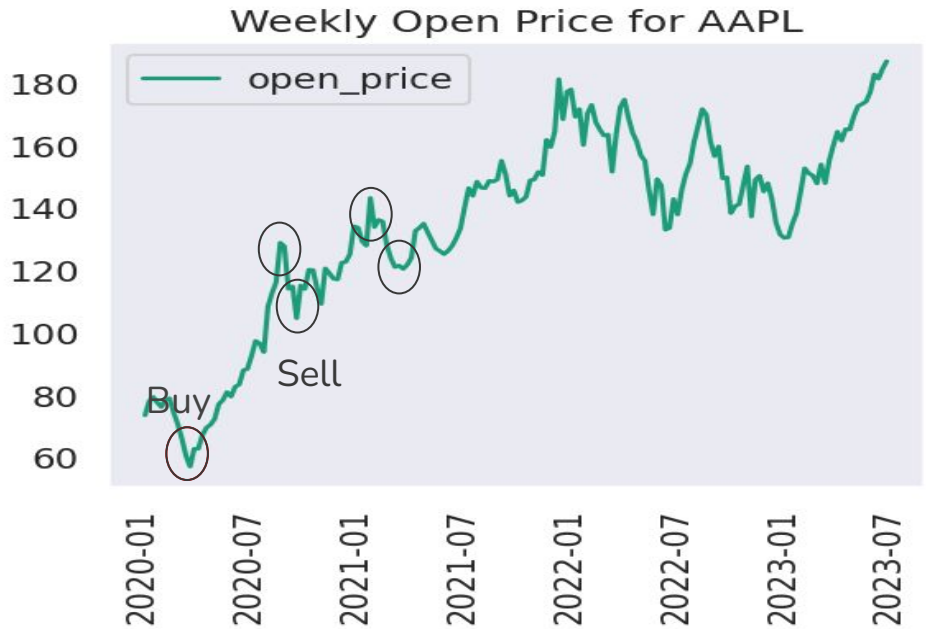
Momentum Trading With S&P 500 Listed US Equities

Using technical indicators to predict weekly increases or decreases in stock prices

Research Question

Can momentum based technical indicators be used to predict whether or not a stock's price will increase or decrease on a weekly basis?



Data

- yfinance python package
  - Weekly equity opening prices, closing prices, and volume from Jan 2020 - June 2023
    - All momentum based features are calculated using this data
- Wikipedia list of S&P 500 companies
  - Scraped using the beautiful soup python package for:
    - List of S&P 500 ticker symbols
    - Industry for each of the S&P 500 companies

	Date	Symbol	Adj Close	Close	High	Low	Open	Volume	ticker	industry
0	2020-01-06	A	85.549835	87.589996	88.239998	83.599998	84.000000	8855200.0	A	Health Care
1	2020-01-13	A	88.020920	90.120003	90.279999	86.699997	87.809998	10250000.0	A	Health Care
2	2020-01-20	A	86.204269	88.260002	90.639999	87.580002	89.800003	6756000.0	A	Health Care
3	2020-01-27	A	80.637001	82.559998	88.360001	82.339996	86.540001	10528800.0	A	Health Care
4	2020-02-03	A	81.232796	83.169998	85.500000	82.110001	83.290001	8224000.0	A	Health Care

Features

All momentum based features are calculated off of the weekly ‘Open’ price of the equities, as well as the weekly volume.

Features Considered:

Industry

Open price

Rolling 2,4,6 week open averages

2,4,6,12 week Exponential Moving Average (EMA)

2,4,6,12 week rolling volume average

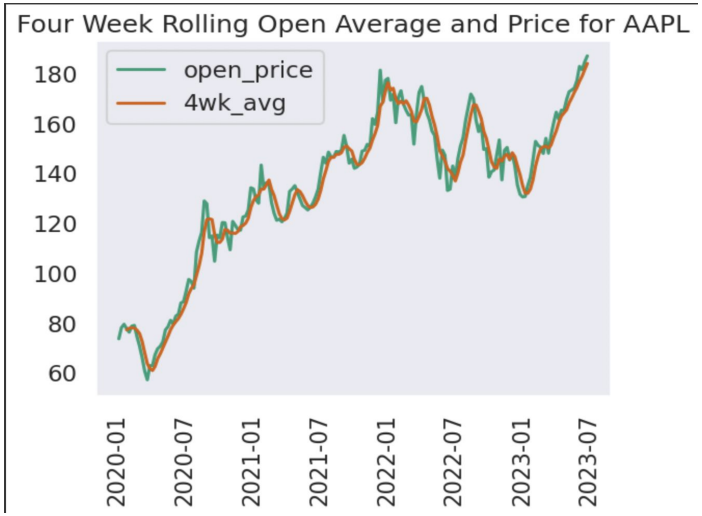
MACD 9-4 week, MACD signal 9-4

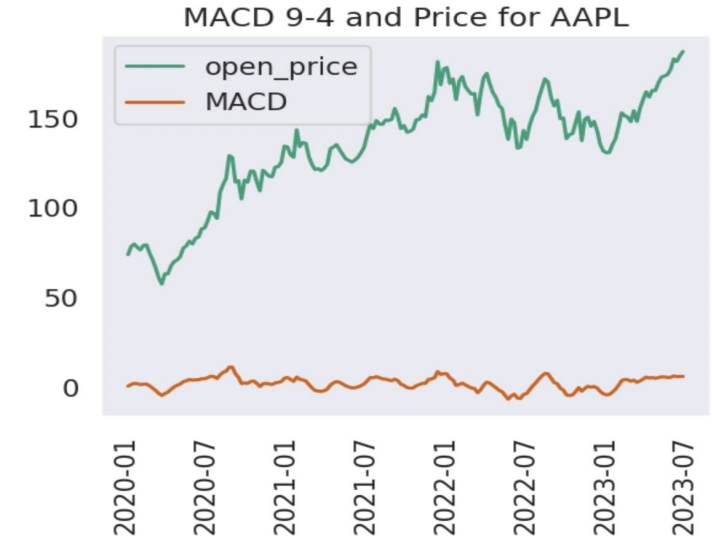
MACD 12-6, MACD signal 12-6

Industry 2,4,6 week open averages

Industry 2,4,6 week volume averages

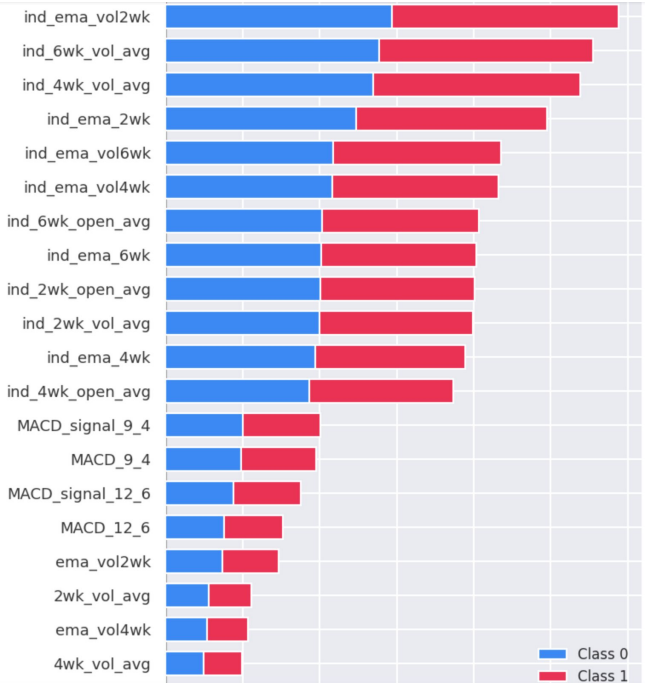
Industry 2,4,6 week EMA





Feature Selection

- A total of 26 momentum based features were considered for model inclusion
- A 27th feature was created by randomly sampling a normal distribution
- All 27 features were fed into a simple random forest model
- Attempt to explain feature importance via SHAP values
- Any feature more ‘important’ than the random variable (which represents pure noise) is a candidate for the final models
- **All features reported as more important than the ‘rv’**



Baseline Model

- The baseline model is an XGBoost classifier
- Default hyperparameters

```
▾                                XGBClassifier
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
```

- Trained on weekly data from Jan 2020 - May 2023
- Tested on data from June 2023
- Target variable == 1 if next week's opening price is greater than the current week's opening price

Metric	**Value**
Accuracy	48.4%
Precision	41.8%
Recall	59.7%

Hyperparameters

- Optimal hyperparameters found via Bayesian Search
  - `gamma': 6.5,`
  - `'learning_rate': 0.15,`
  - `'max_depth': 14,`
  - `'n_estimators': 80,`
  - `'reg_alpha': 0.8,`
  - `'reg_lambda': 0.1`

Results - Optimized Model

Metric	**Value**
Accuracy	45.5%
Precision	41%
Recall	65%

Discussion

- The proposed model and method did **not** yield a viable trading strategy
- Weekly price movements seem to be more complicated than what can be measured by moving averages and other technical indicators.
- SHAP values from the optimized model suggest that industry specific trends are more predictive of price movement than company specific trends in the short term
- More industry specific features AND company specific features would likely improve the model
- Future work to include news articles for sentiment or topic modeling
- Different models (LSTM)