XAI models applied to investing

CentraleSupélec

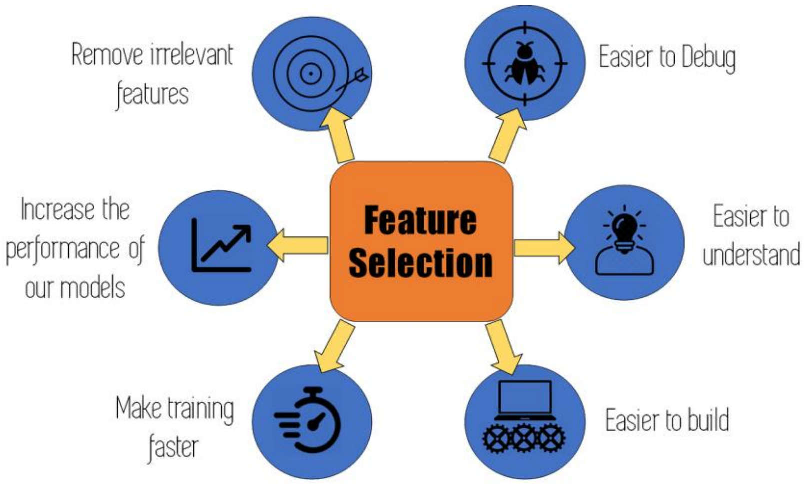
Alban Sarfati

https://github.com/albnsft/XAI_Investing

Abstract

- Machine learning techniques have recently become the norm for detecting patterns in financial markets
- Investing involves making decisions about buying or selling securities based on various factors
- Relying solely on machine learning algorithms to make decisions can have negative consequences
- Can an investor explain clearly the intuition of his position on securities ?

Introduction



SHAP

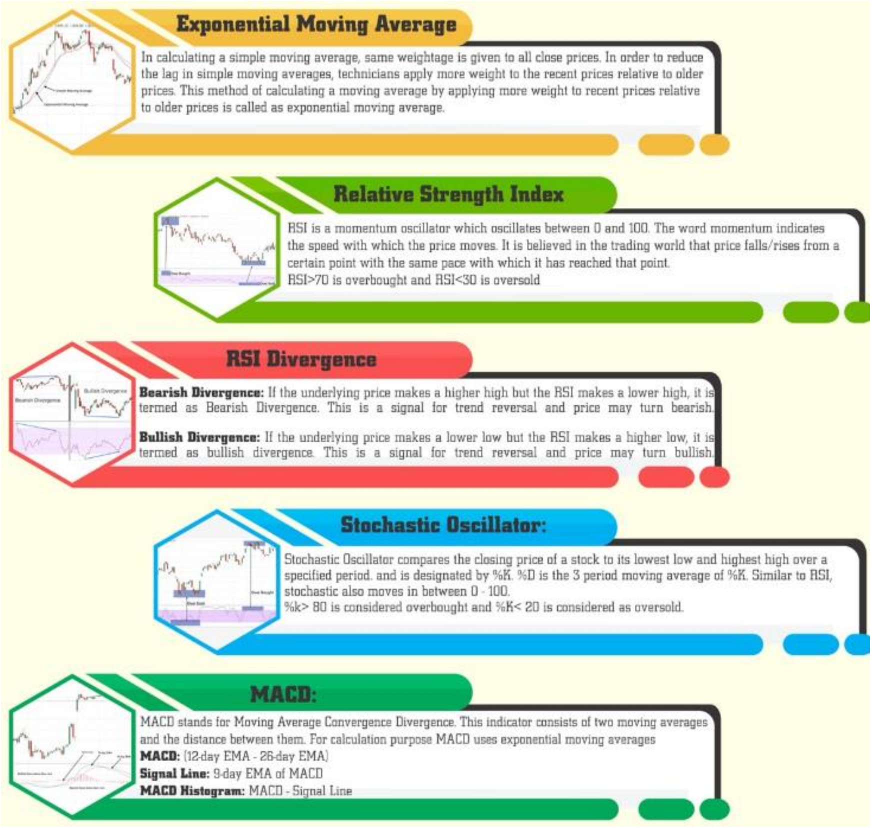
- A basis investment task it to make prediction about the future direction of securities movements
- The goal in this work is to use ”black boxes” model to balance accuracy and eXplainability
- Tailoring a feature selection framework that increases the accuracy of the employed ML model
- Using the SHAP values to evaluate the contributions of each individual feature to the direction

Data

	High	Low	Close	Volume
2023-02-13 00:00:00	154.25999	150.92000	153.85001	62199000.00000
2023-02-14 00:00:00	153.77000	150.86000	153.20000	61707600.00000
2023-02-15 00:00:00	155.50000	152.88000	155.33000	65669300.00000
2023-02-16 00:00:00	156.33000	153.35001	153.71001	68167900.00000
2023-02-17 00:00:00	153.00000	150.85001	152.55000	59095900.00000
2023-02-21 00:00:00	151.30000	148.41000	148.48000	58867200.00000
2023-02-22 00:00:00	149.95000	147.16000	148.91000	51011300.00000
2023-02-23 00:00:00	150.34000	147.24001	149.39999	48349600.00000

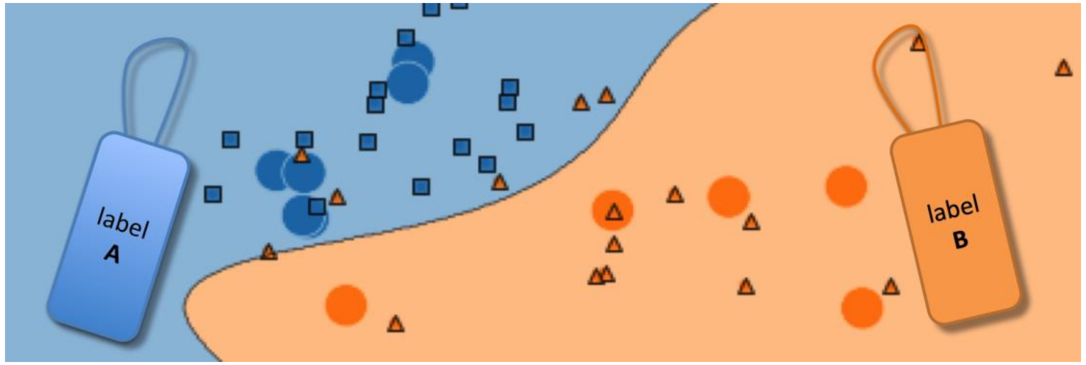
- Apple and Google stocks daily historical data from yahoo finance over the last 20 years
- Each instance is indexed by a timestamp and characterized by the open price, close price, highest price, lowest price and volume
- Based on this information, for each stock, the features and labels have been engineered allowing to form the training, validation and testing dataset

Features



- Technical indicators represent statistical tools generating signals
- Using eight of them to identify trends, regime switches, momentum and potential reversal points in the stock market
- By combining these features with different time windows: total of 17 features

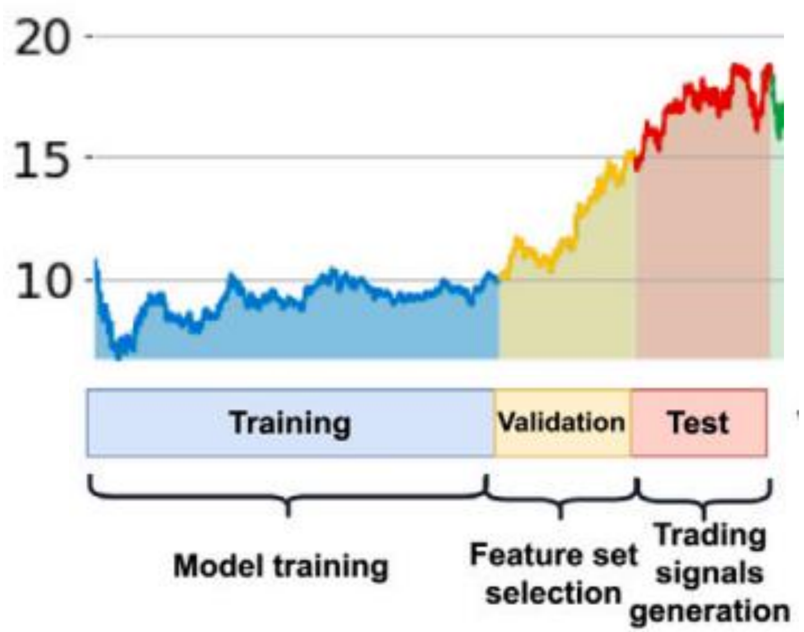
Label



- The label is a binary variable given by the price direction
- Used to predict upward or downward movements in the market
- The direction's predictions of the model are transformed into investment positions

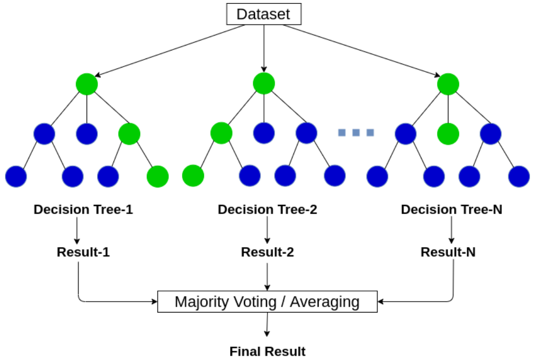
Environment

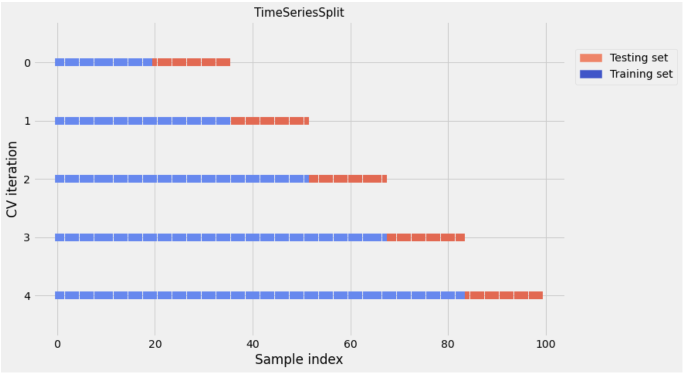
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$



- Features are scaled within the range (0, 1)
- Training split corresponds to the first 70% time series instances, while validation and testing are 15% (2017 to 2019 and 2020 to 2023)
- The validation part is used for the feature selection process
- The testing part evaluate if this selection has improved the performance

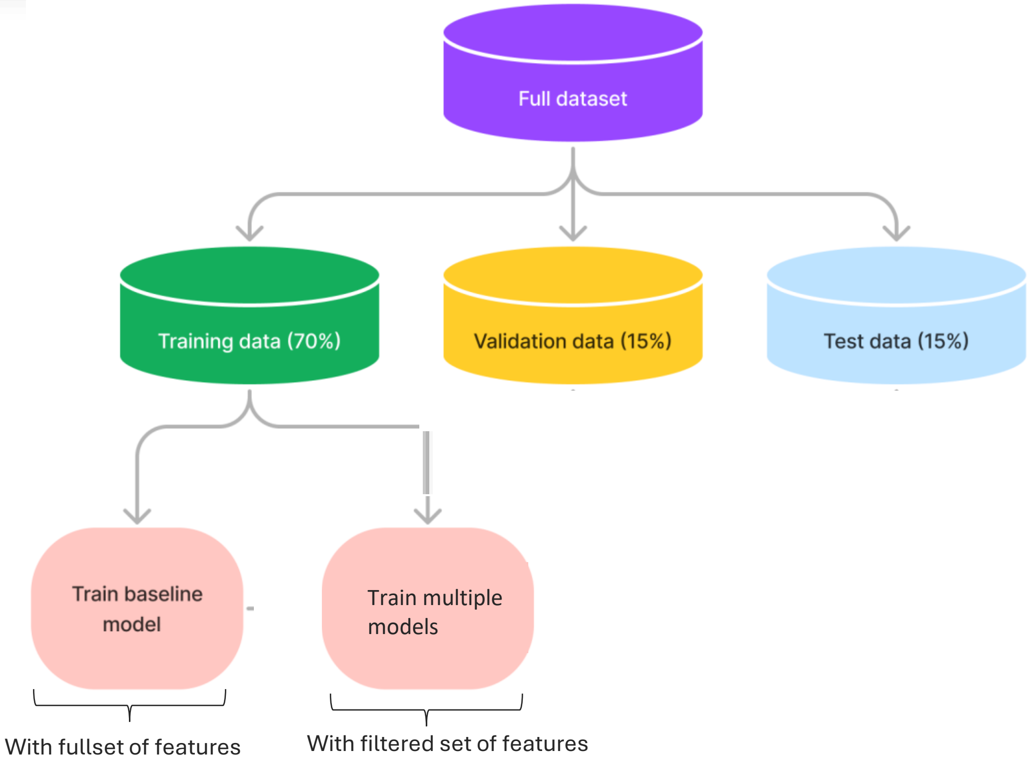
Model





- Random Forests can handle noisy and outliers instances
- Tuning the hyper-parameters:
  (i) The number of trees in the forest
  (ii) The maximum depth of each tree
  (iii) The maximum number of features
- By using :
  (i) Cross validation approach (TimeSeriesSplit) on only the training set
  (ii) Full set of features
  (iii) Maximizing accuracy score

Training



- Based on the optimal hyper-parameters, multiple RandomForestClassifier have been trained using either the full set of features or only filtered ones
- The model associated with the full set of features is the baseline and will be the benchmark for performance comparisons

Metrics

Trading Strategy :

*At each time step t, the return of the strategy is $position_{t-1} * stockreturn_t$
*Deduction of transaction costs when a trade has taken place
*Calculation of the net asset value of the strategy and market (naive)
*Calculation of the absolute and the out performance of the strategy

- The performance of these models are judged according to different metrics:
- Accuracy, f1-score, AUC, brier-score
- The good performance of these metrics is not necessarily correlated to a good financial performance
- Important to correctly predict large market movements and not just the majority of market movements

Features selection

Impurity importance

Permutation importance

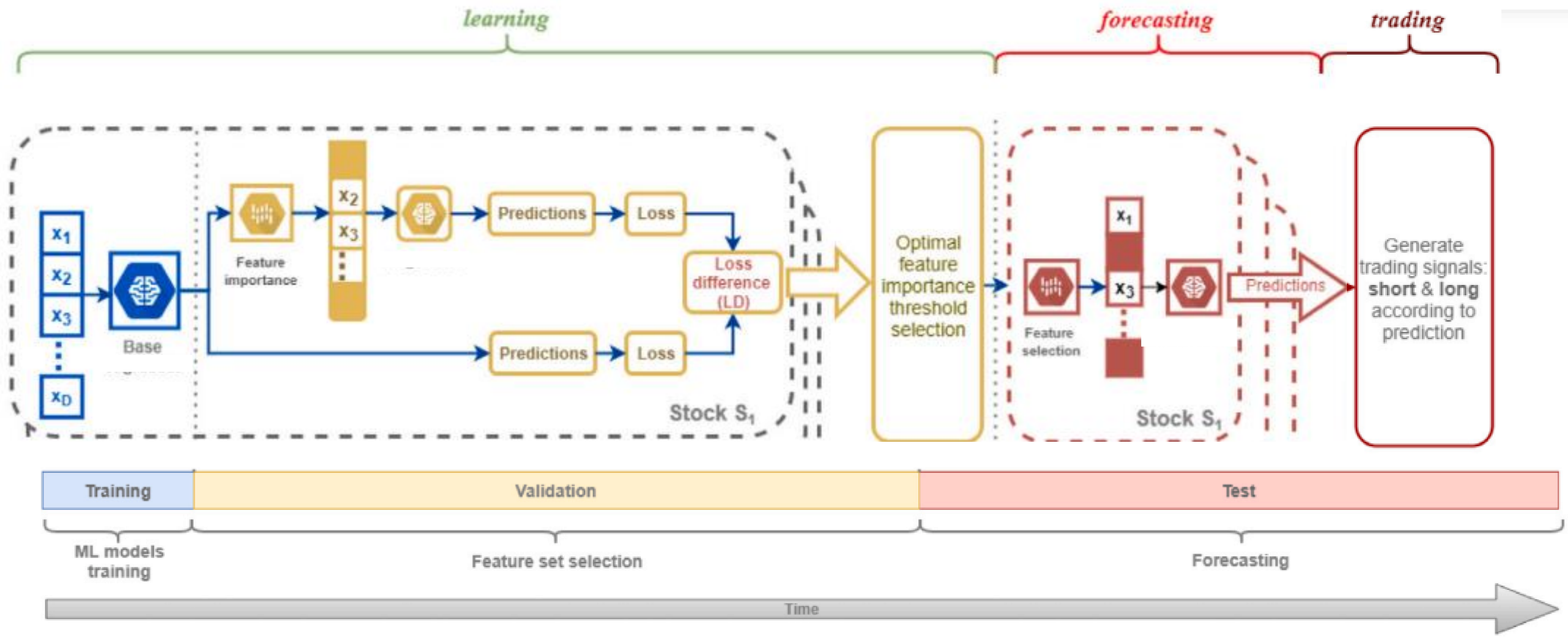
Lime importance

SHAP importance

- Assigning importance scores to features
- The methods' common properties are :
- Model agnostic
- Computes the feature importance on validation set
- Has no tuning parameters

Removal Strategy

Goal : Looking for the combination of removable features that leads to the largest increase in accuracy over the validation set



Results - Apple

Method	accuracy	f1-score	auc	brier	outperformance	removed features
baseline	0.46	0.4	0.5	0.26	-1.15	None
impurity	0.46	0.42	0.51	0.25	-1.29	[Direction3, Direction21, Direction1]
permutation	0.56	0.49	0.5	0.25	0.73	[Direction21, EWMA7, EWMA14, EWMA21, Volatility14]
lime	0.48	0.47	0.5	0.25	-1.21	[EWMA7, Direction3, Direction21, Direction2, Direction1]
shap	0.46	0.42	0.51	0.25	-1.29	[Direction3, Direction1, Direction21]

Table 1. Apple validation metrics for feature selection methods.

Method	accuracy	f1-score	auc	brier	outperformance	removed features
baseline	0.49	0.45	0.52	0.25	-1.09	None
impurity	0.51	0.48	0.54	0.25	-0.47	[Direction3, Direction21, Direction1]
permutation	0.53	0.44	0.54	0.25	1.65	[Direction21, EWMA7, EWMA14, EWMA21, Volatility14]
lime	0.51	0.51	0.53	0.25	-0.81	[EWMA7, Direction3, Direction21, Direction2, Direction1]
shap	0.51	0.48	0.54	0.25	-0.47	[Direction3, Direction1, Direction21]

Table 2. Apple testing metrics for feature selection methods.

Results - Google

Method	accuracy	f1-score	auc	brier	outperformance	removed features
baseline	0.51	0.45	0.48	0.25	0.05	None
impurity	0.53	0.44	0.5	0.25	0.08	[Direction2, Direction3, Direction5, Direction21]
permutation	0.57	0.56	0.56	0.25	2.16	[OBV, EWMA7, Direction3, RSI, Direction1, ..]
lime	0.54	0.4	0.51	0.25	0.15	[EWMA7, Direction5, Direction21]
shap	0.53	0.44	0.5	0.25	0.08	[Direction1, Direction2, Direction3, Direction21]

Table 3. Google validation metrics for feature selection methods.

Method	accuracy	f1-score	auc	brier	outperformance	removed features
baseline	0.52	0.43	0.53	0.25	-0.25	None
impurity	0.53	0.42	0.5	0.25	-0.01	[Direction2, Direction3, Direction5, Direction21]
permutation	0.53	0.53	0.51	0.25	1.01	[OBV, EWMA7, Direction3, RSI, Direction1, ..]
lime	0.53	0.38	0.52	0.25	0.01	[EWMA7, Direction5, Direction21]
shap	0.53	0.42	0.5	0.25	-0.01	[Direction1, Direction2, Direction3, Direction21]

Table 4. Google testing metrics for feature selection methods.

→ Introduce clearly improvements both in terms of predictive and financial performance
→ Permutation method is the one working the best

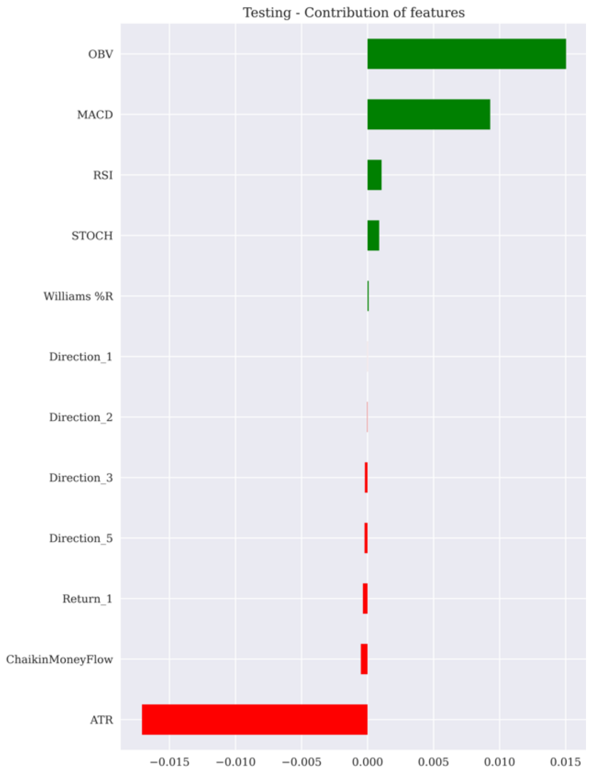
Trading Strategy

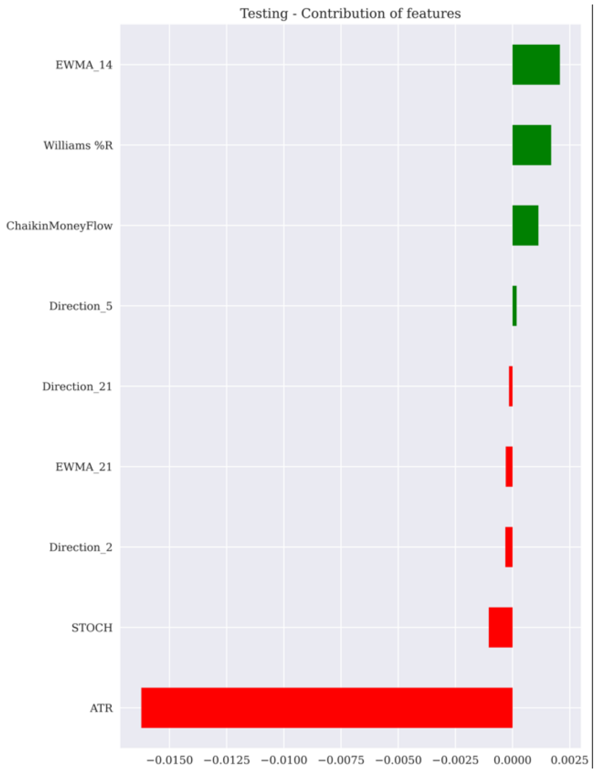
Table 5. Apple testing trading strategy with permutation method.

Table 6. Google testing trading strategy with permutation method.

→ Providing long-short predictive signals whose information content suffices and is less noisy than the one embedded in the full set of features.

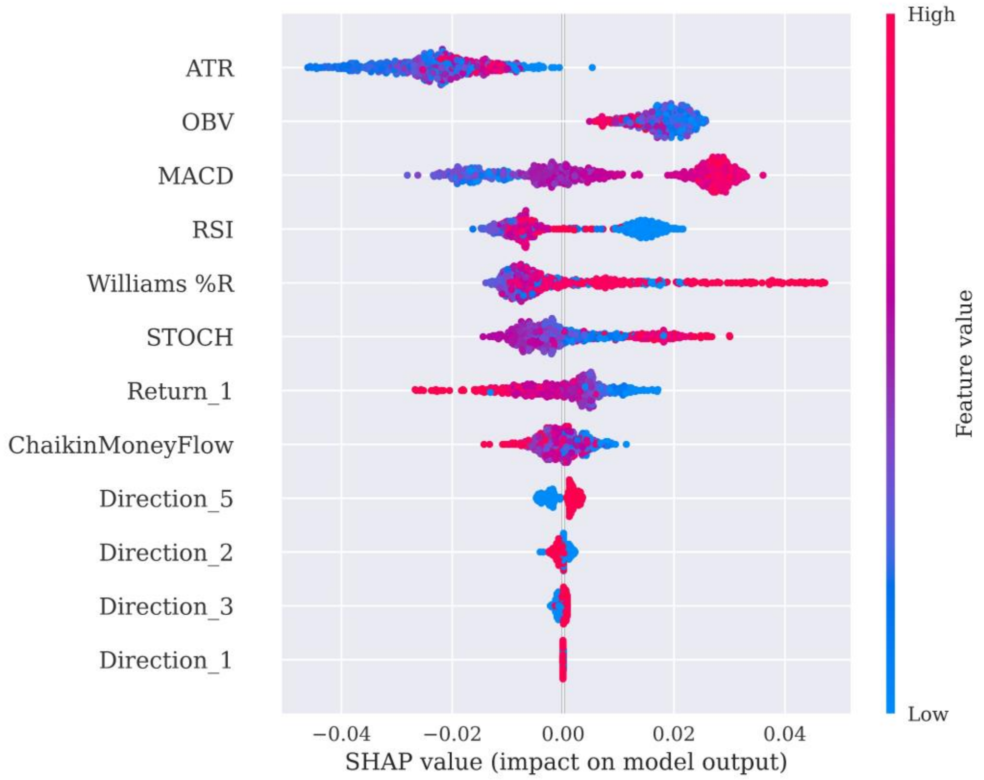
Features interpretation





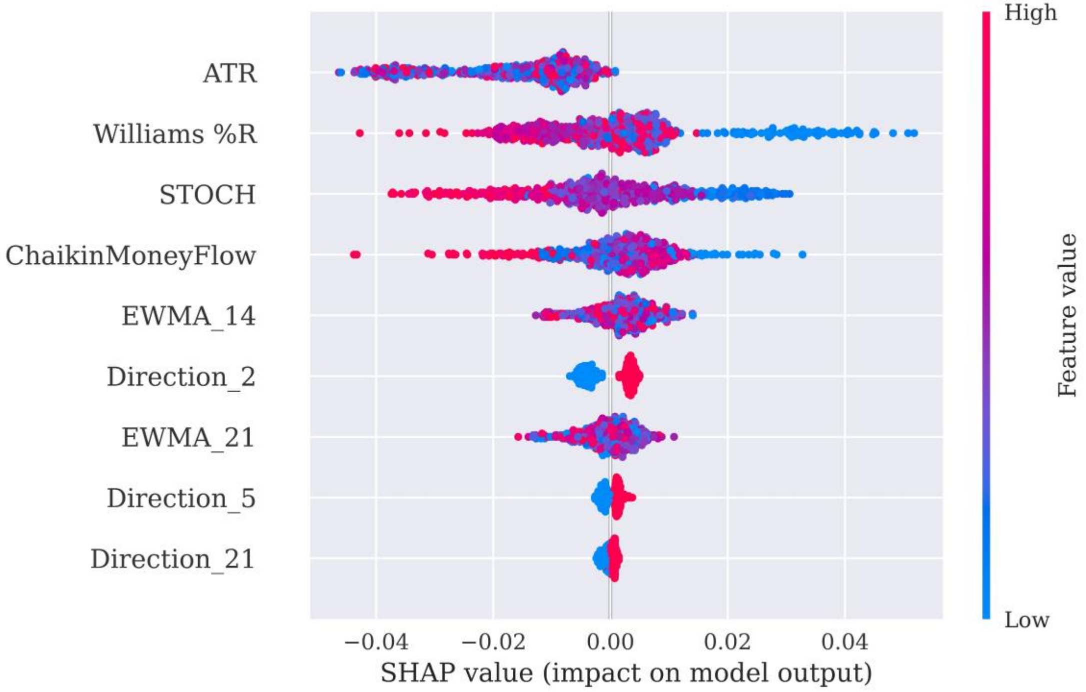
- Using Shapley values
- The testing predictions are analysed
- We can infer the contribution of each feature to the upward or downward movement
- Some Shapley value don’t reflect the true contribution, as max accuracy = 0.53

Apple Summary plot



- Direction features seem to be the least significant
- Most significant feature to a downward movement is ATR which measures market volatility
- MACD has a linear trend with the upward probability direction

Google Summary plot



- The higher the values of Williams%R, STOCK and ChaikinMoneyFlow the higher they contribute to a downward probability direction
- Shapley values allow for a global understanding of the model behaviour
- Has traits in common with the trading decisions of an investor relying on the literature of technical indicators and market factors

Conclusion

- Automatic feature selection is a possible solution to promote greater reliability and robustness in explainable artificial intelligence
- Importance of understanding the global predictions of AI models through feature's contribution to the upward probability
- Must be careful not becoming overconfident about the forecasting ability of the model as it is generating often false positive and false negative signals.