European Union
European Regional Development Fund

EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη
Partnership Agreement 2014 - 2020

Deep Learning for Financial Trading using Sentiment Analysis

A. Tsantekidis[1], N. Passalis[1], P. Nousi[1], S. Doropoulos[2], A. Fanakis[2] and A. Tefas[1]

Computational Intelligence and Deep Learning (CIDL) group,
AIIA Lab., Dept. of Informatics,
Aristotle University of Thessaloniki[1]
https://cidl.csd.auth.gr/

DataScouting[2]
https://datascouting.com/

DeepFinance consortium

Academic Partner

ARISTOTLE UNIVERSITY OF THESSALONIKI

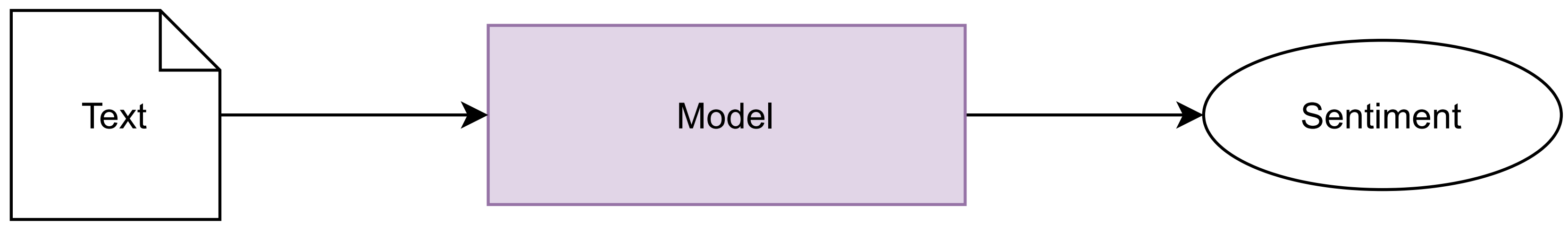
Industrial Partners

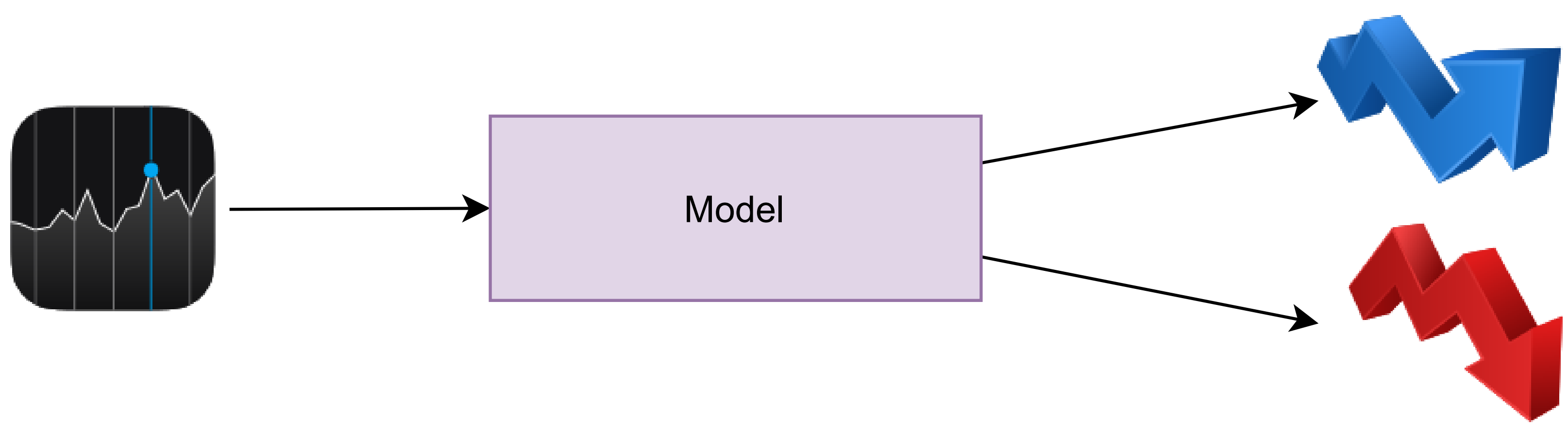
DATASCOUTING

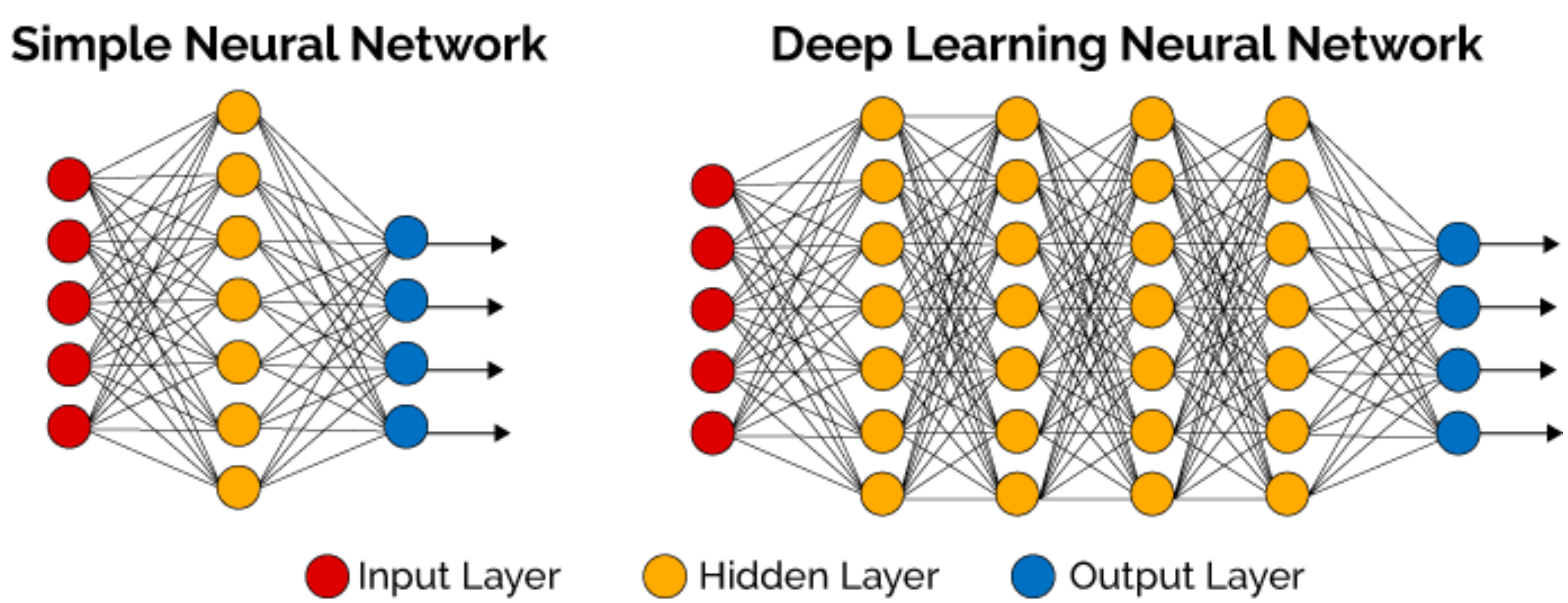
SPEEDLAB

European Union
European Regional Development Fund
EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION
ΕΣΠΑ
2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

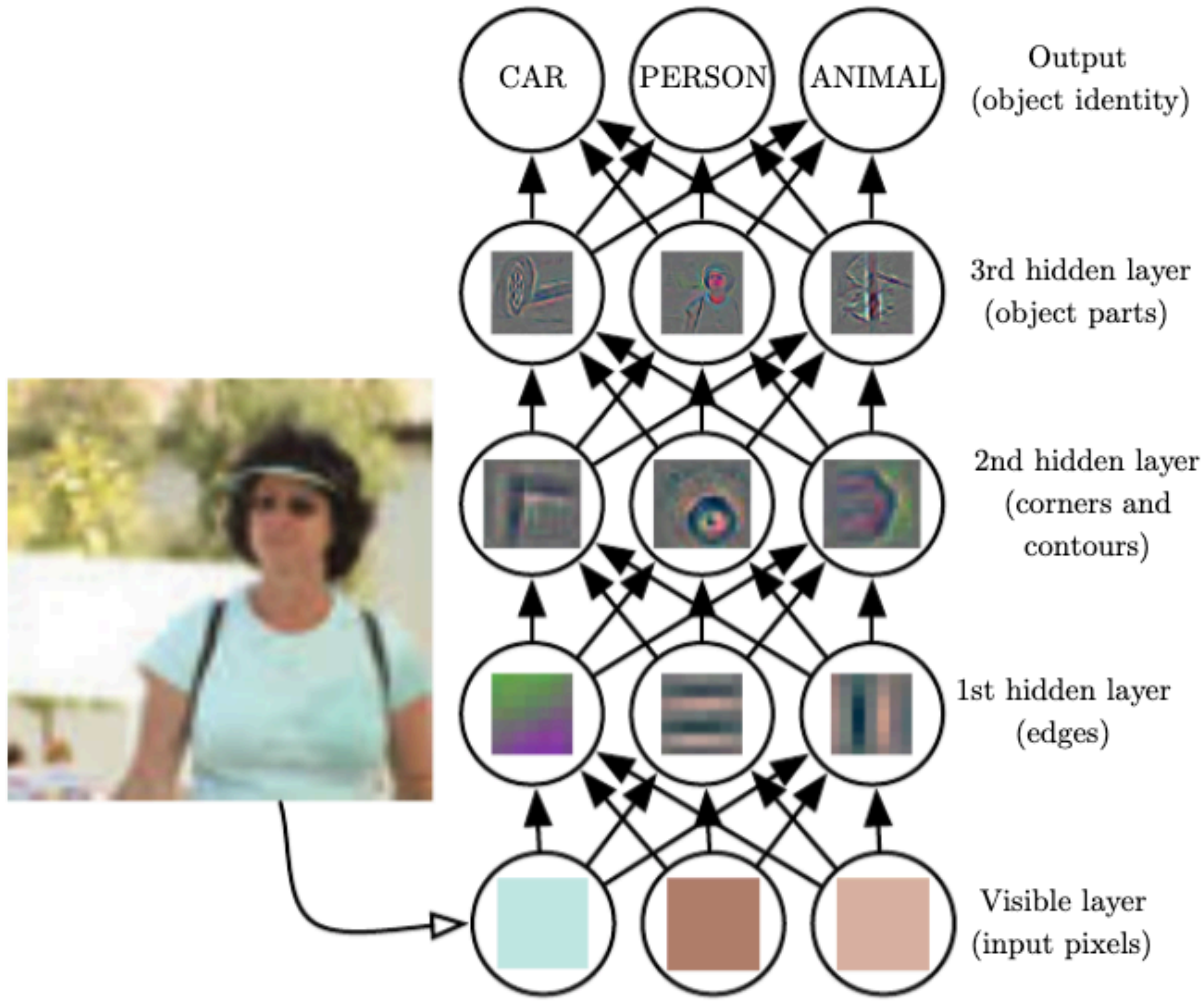
Deep Learning



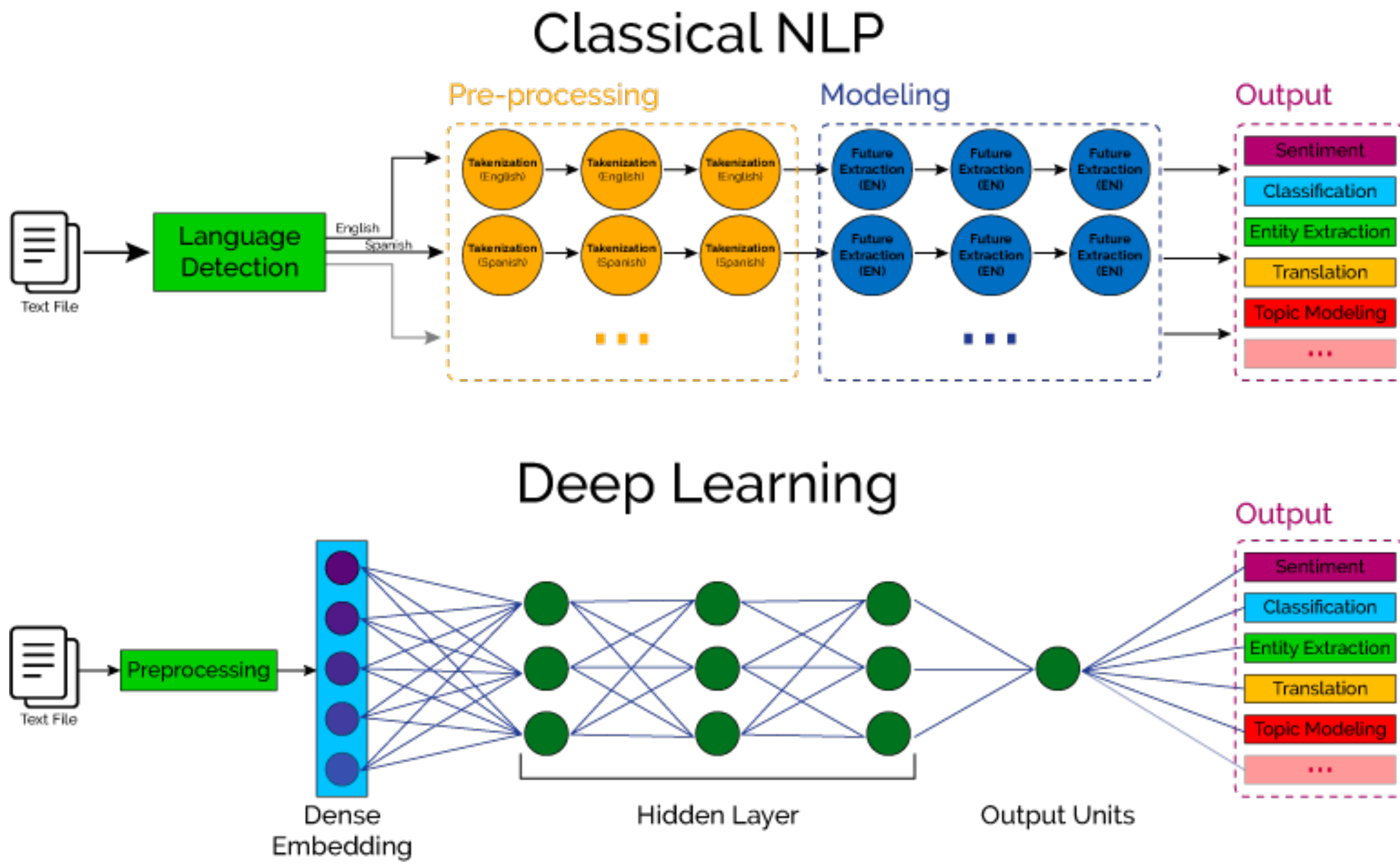


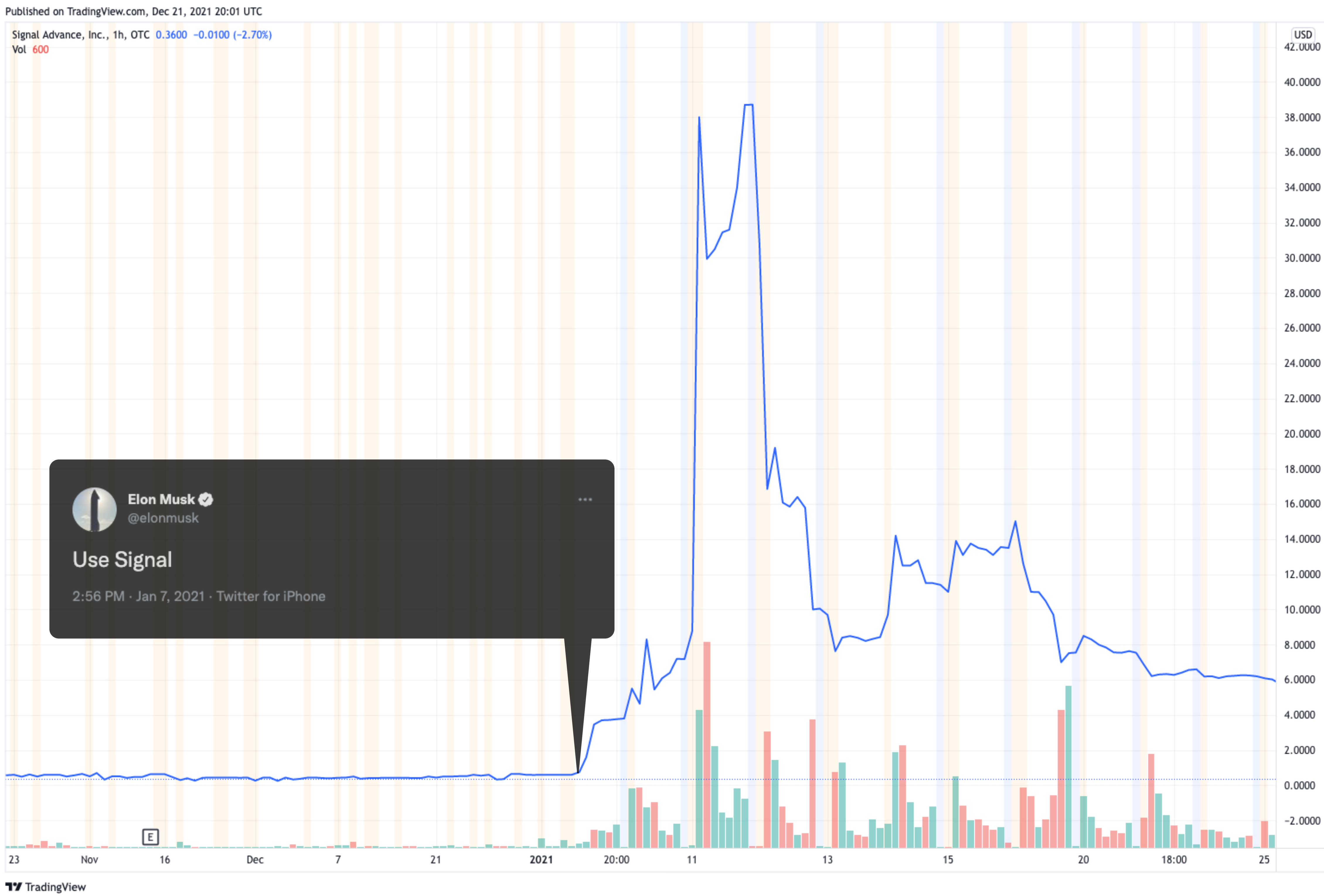


Deep Learning



Deep Learning





Published on TradingView.com, Dec 21, 2021 20:01 UTC
Signal Advance, Inc., 1h, OTC 0.3600 −0.0100 (−2.70%)
Vol 600
USD
42.0000
40.0000
38.0000
36.0000
34.0000
32.0000
30.0000
28.0000
26.0000
24.0000
22.0000
20.0000
18.0000
16.0000
14.0000
12.0000
10.0000
8.0000
6.0000
4.0000
2.0000
0.0000
−2.0000
Elon Musk
@elonmusk
Use Signal
2:56 PM · Jan 7, 2021 · Twitter for iPhone
E
23
Nov
16
Dec
7
21
2021
20:00
11
13
15
20
18:00
25
TradingView


Elon Musk @elonmusk · Dec 20, 2020
One word: Doge
10.6K
31.8K
215.6K
0.0100
0.0090
0.0080
0.0070
0.0060
0.0050
0.0040
0.0030
0.0020
Dec
7
13
19
25
2021
7
13
19
25
CoinMarketCap

Algorithmic Trading

- Optimal Execution
- Market Making
- Statistical Arbitrage
    - Reduce exposure to markets
    - Leveraging repeatable patterns

Statistical Arbitrage

- Strategies based on short-term and large number of trades.

- Uses price time-series, such as the Open-High-Low-Close to extract Technical Indicators.

Statistical Arbitrage

- Strategies based on short-term and large number of trades.

- Uses price time-series, such as the Open-High-Low-Close to extract Technical Indicators.

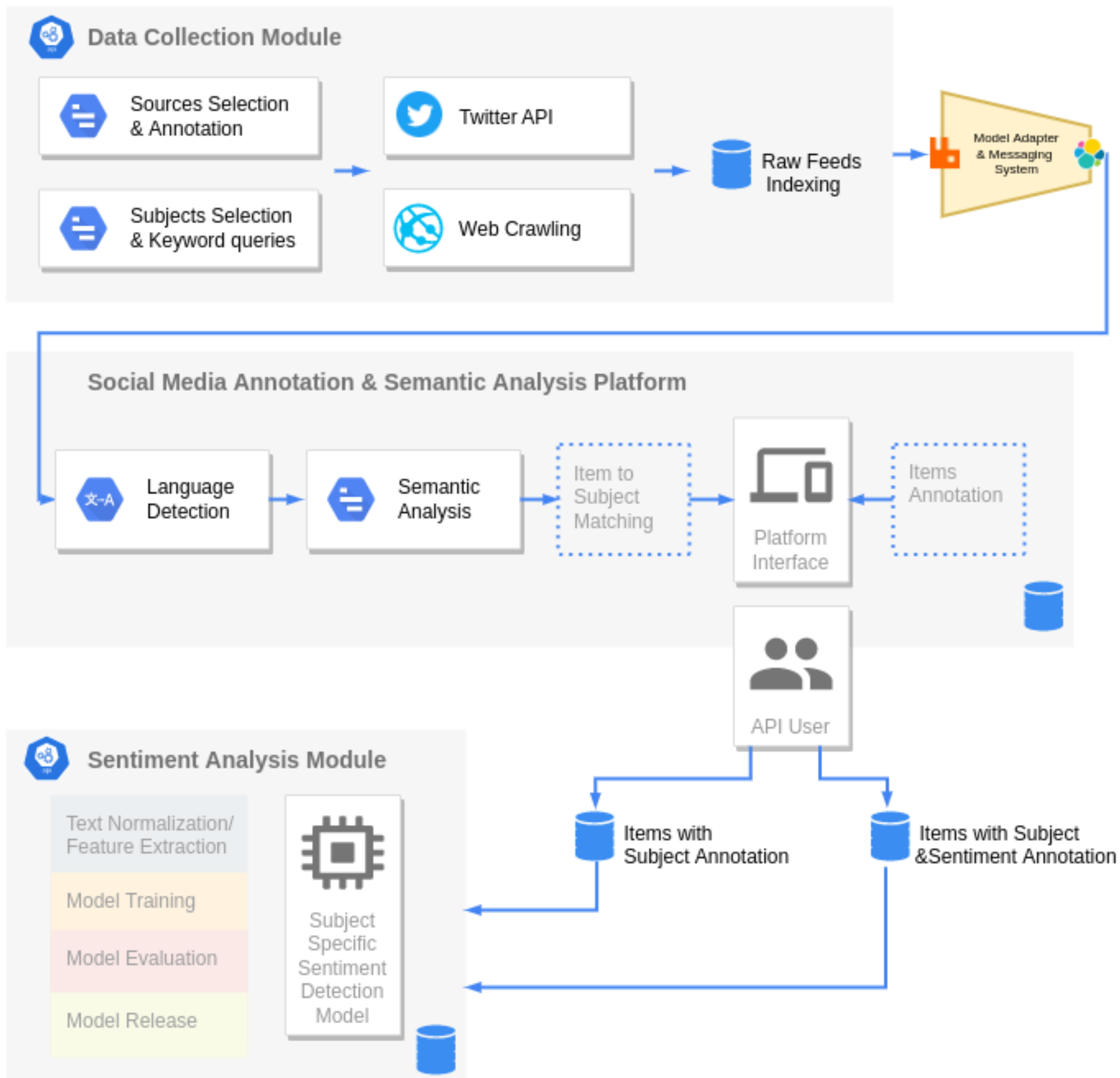
Statistical Arbitrage

- Strategies based on short-term and large number of trades.
- Uses price time-series, such as the Open-High-Low-Close to extract Technical Indicators.

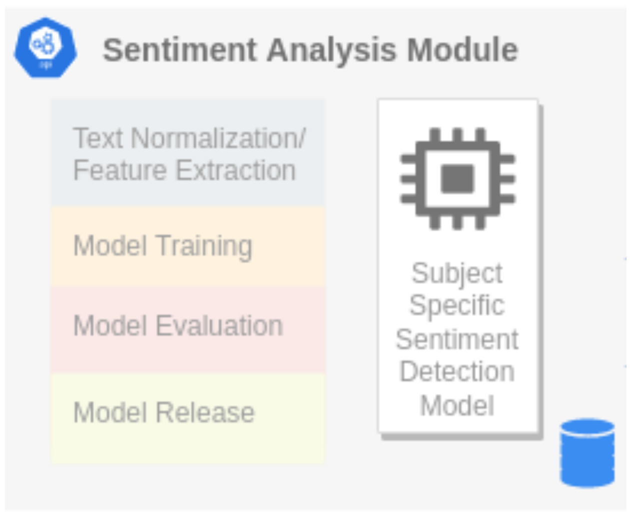
Sentiment

- Social media (Twitter, Facebook, etc.)
- News articles, Blogs
- Public forums / chat groups

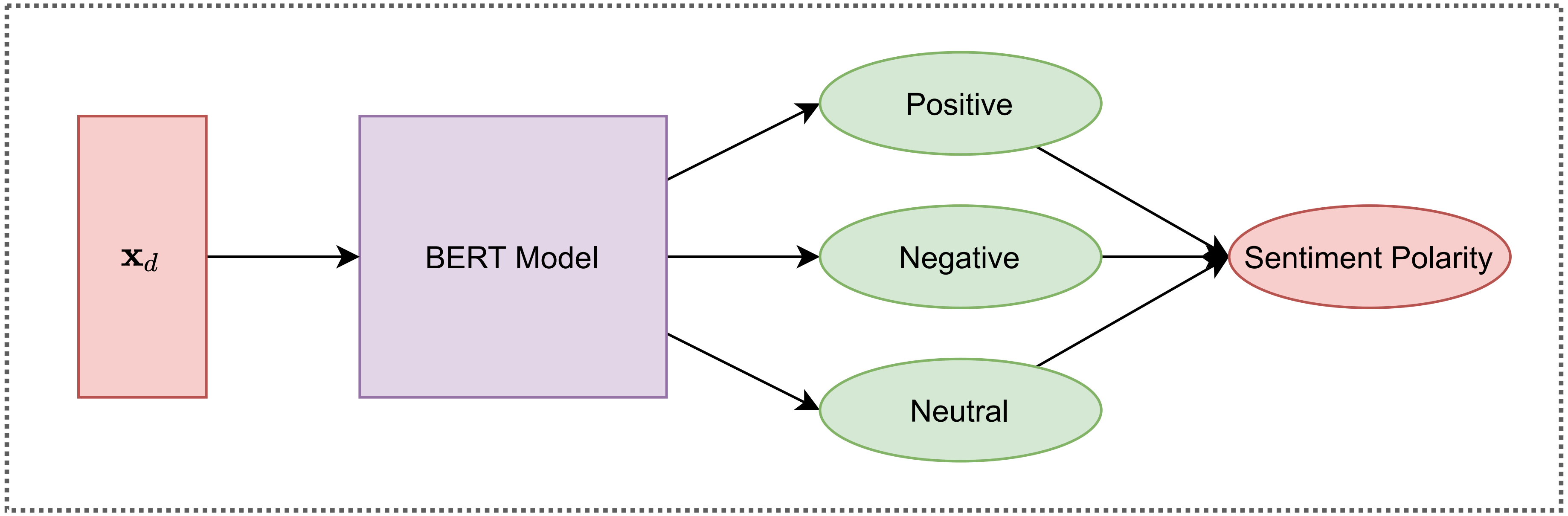
Dataset Creation



Dataset Creation



Sentiment Extractor



Dataset

Dataset

Price Time-Series Features

$$\mathbf{x}_t^p = \left[\frac{c_{t-L-1}}{c_{t-L-2}} - 1, \ldots, \frac{c_t}{c_{t-1}} - 1\right] \in \mathbb{R}^L$$

Dataset

Price Time-Series Features

$$\mathbf{x}_t^p = \left[\frac{c_{t-L-1}}{c_{t-L-2}} - 1, \ldots, \frac{c_t}{c_{t-1}} - 1\right] \in \mathbb{R}^L$$

Sentiment Polarity Vector

$$x_t^s = \frac{1}{|\mathcal{X}_t|} \sum_{\mathbf{x}_d \in \mathcal{X}_t} f\left(\mathbf{x}_d\right) \in \mathbb{R}^3$$

$$\mathbf{x}_t^s = \left[x_{t-L-1}^s, \ldots, x_t^s\right] \in \mathbb{R}^{L \times 3}$$

Fused Tensor

$$\mathbf{x}_t = \left[\mathbf{x}_t^p; \mathbf{x}_t^s\right] \in \mathbb{R}^{L \times 5}$$

Dataset

Price Time-Series Features

$$\mathbf{x}_t^p = \left[\frac{c_{t-L-1}}{c_{t-L-2}} - 1, \ldots, \frac{c_t}{c_{t-1}} - 1\right] \in \mathbb{R}^L$$

Sentiment Polarity Vector

$$x_t^s = \frac{1}{|\mathcal{X}_t|} \sum_{\mathbf{x}_d \in \mathcal{X}_t} f\left(\mathbf{x}_d\right) \in \mathbb{R}^3$$

$$\mathbf{x}_t^s = \left[x_{t-L-1}^s, \ldots, x_t^s\right] \in \mathbb{R}^{L \times 3}$$

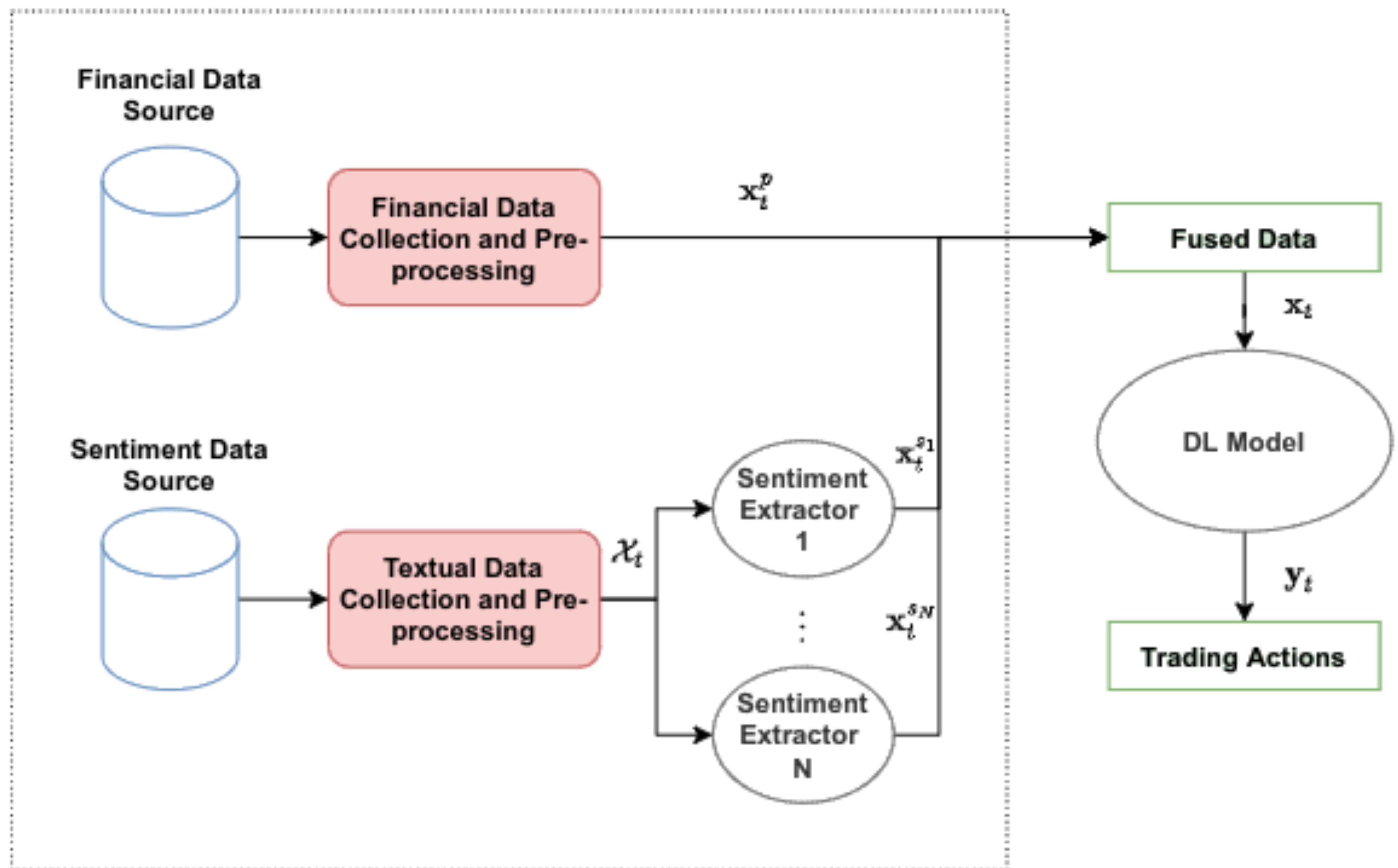
Fused Tensor

$$\mathbf{x}_t = \left[\mathbf{x}_t^p; \mathbf{x}_t^s\right] \in \mathbb{R}^{L \times 5}$$

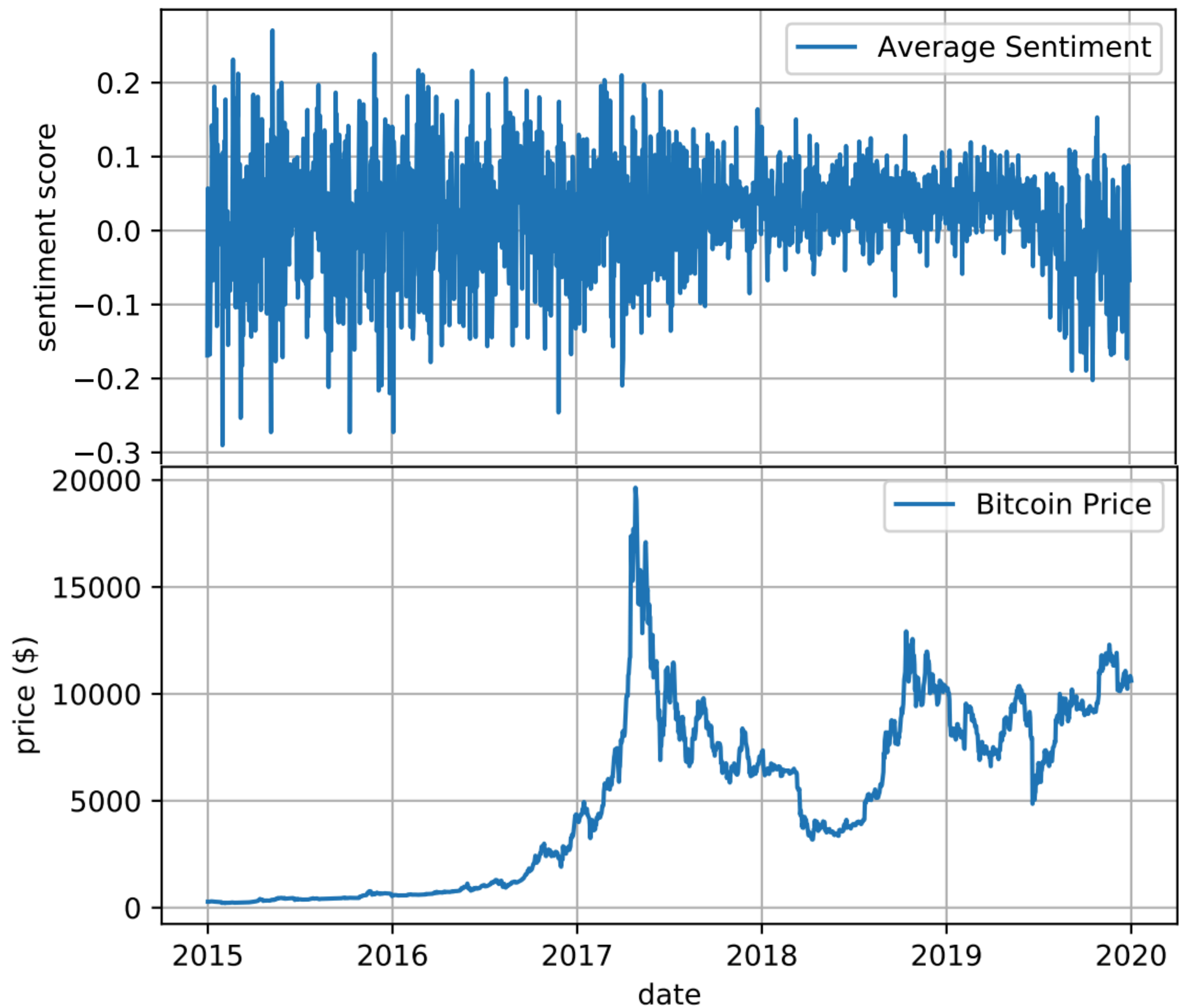
Target Labels

$$l_t = \begin{cases} 1 & \text{if } \frac{c_{t+1}}{c_t} - 1 > c_{\text{thres}} \\ -1 & \text{if } \frac{c_{t+1}}{c_t} - 1 < -c_{\text{thres}} \\ 0 & \text{otherwise} \end{cases}$$

Forecasting Pipeline

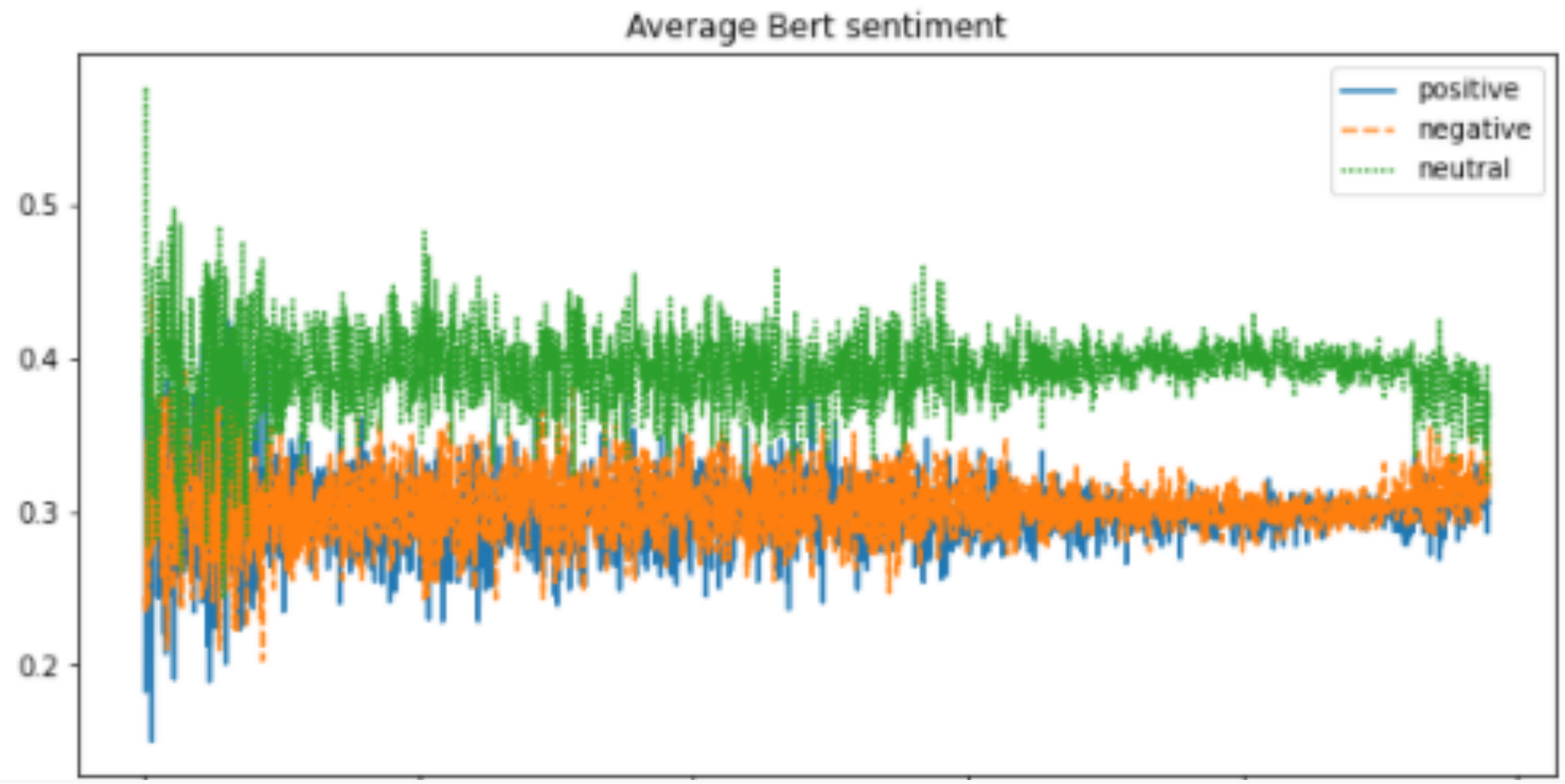


Results

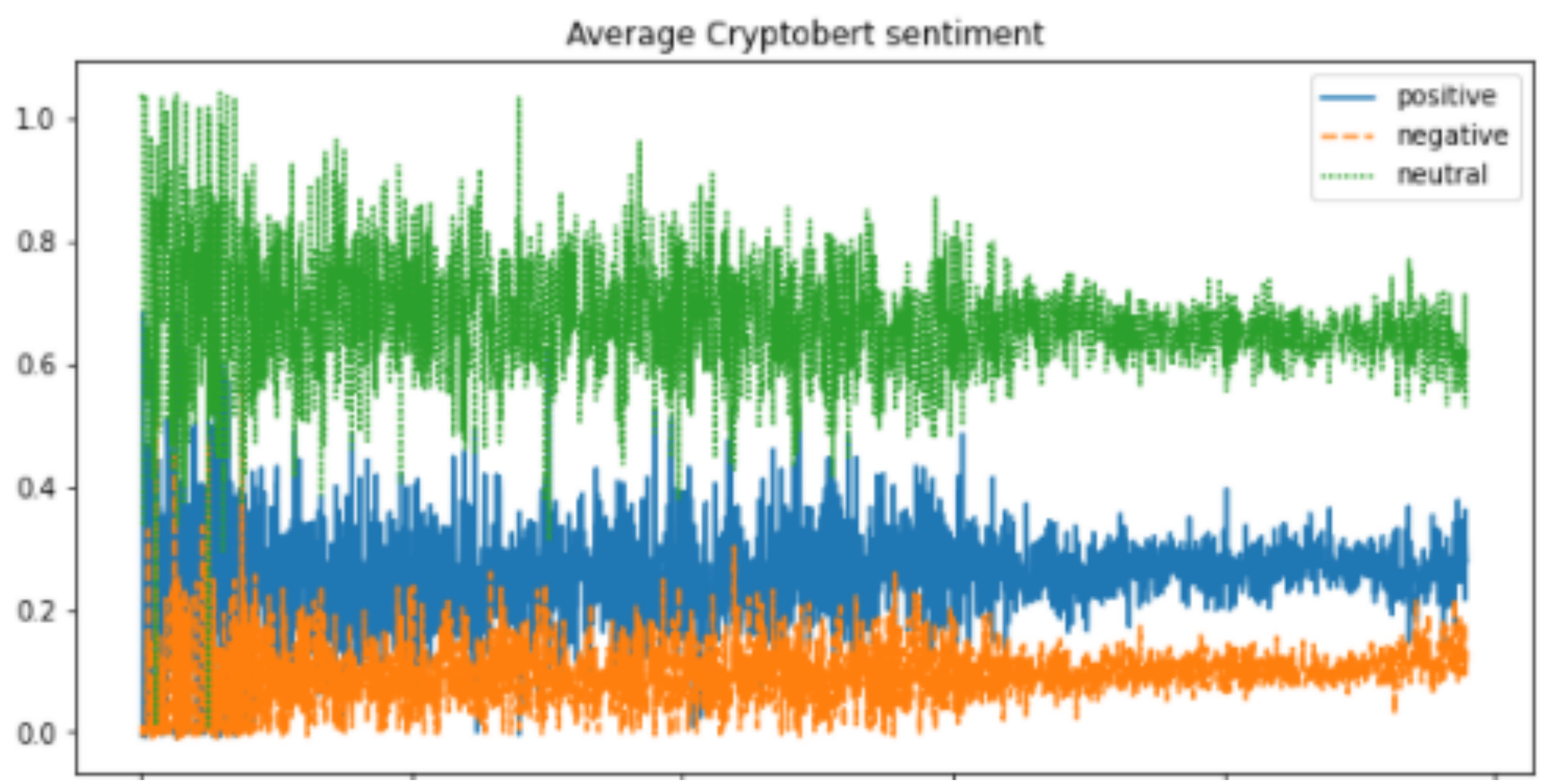


Average Sentiment
sentiment score
0.2
0.1
0.0
−0.1
−0.2
−0.3
Bitcoin Price
price ($)
20000
15000
10000
5000
0
2015
2016
2017
2018
2019
2020
date

Results



Average Bert sentiment
positive
negative
neutral
0.5
0.4
0.3
0.2



Average Cryptobert sentiment
positive
negative
neutral
1.0
0.8
0.6
0.4
0.2
0.0

Results

	BERT (Acc.)	**BERT (PnL)**	**CryptoBERT (Acc.)**	**CryptoBERT (PnL)**
MLP (s)	$45.1\% \pm 2.7\%$	$106.8\% \pm 22.8\%$	$43.6\% \pm 4.4\%$	$\mathbf{186.6}\% \pm \mathbf{24.5}\%$
MLP (s + p)	$\mathbf{45.9}\% \pm \mathbf{3.4}\%$	$\mathbf{164.0}\% \pm \mathbf{20.6}\%$	$\mathbf{45.3}\% \pm \mathbf{3.3}\%$	$161.2\% \pm 20.6\%$
CNN (s)	$44.1\% \pm 4.6\%$	$\mathbf{200.7}\% \pm \mathbf{17.7}\%$	$43.8\% \pm 4.6\%$	$\mathbf{197.2}\% \pm \mathbf{18.7}\%$
CNN (s + p)	$\mathbf{47.3}\% \pm \mathbf{2.9}\%$	$177.0\% \pm 18.7\%$	$\mathbf{47.3}\% \pm \mathbf{2.8}\%$	$173.7\% \pm 16.2\%$
LSTM (s)	$44.5\% \pm 3.1\%$	$127.5\% \pm 43.7\%$	$43.4\% \pm 3.8\%$	$169.1\% \pm 27.6\%$
LSTM (s + p)	$\mathbf{51.7}\% \pm \mathbf{2.6}\%$	$\mathbf{231.3}\% \pm \mathbf{22.1}\%$	$\mathbf{51.5}\% \pm \mathbf{2.8}\%$	$\mathbf{222.1}\% \pm \mathbf{18.9}\%$

	Test	**PnL**
MLP (s)	$43.9\% \pm 4.9\%$	$\mathbf{182.6}\% \pm \mathbf{24.5}\%$
MLP (s + p)	$\mathbf{46.6}\% \pm \mathbf{3.0}\%$	$176.2\% \pm 20.4\%$
CNN (s)	$44.2\% \pm 4.4\%$	$\mathbf{203.2}\% \pm \mathbf{23.6}\%$
CNN (s + p)	$\mathbf{47.1}\% \pm \mathbf{2.7}\%$	$181\% \pm 21.2\%$
LSTM (s)	$43.7\% \pm 3.9\%$	$173.4\% \pm 21\%$
LSTM (s + p)	$\mathbf{52.2}\% \pm \mathbf{3}\%$	$\mathbf{225.1}\% \pm \mathbf{22.4}\%$

Conclusion

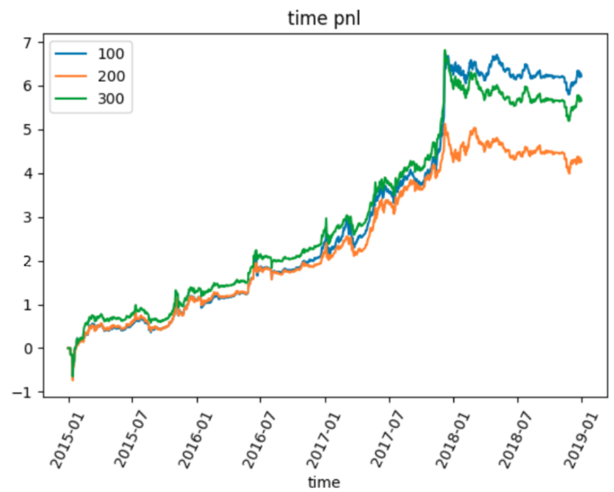
- Sentiment data from online source (social media, news, etc.) can be used to augment trading systems.

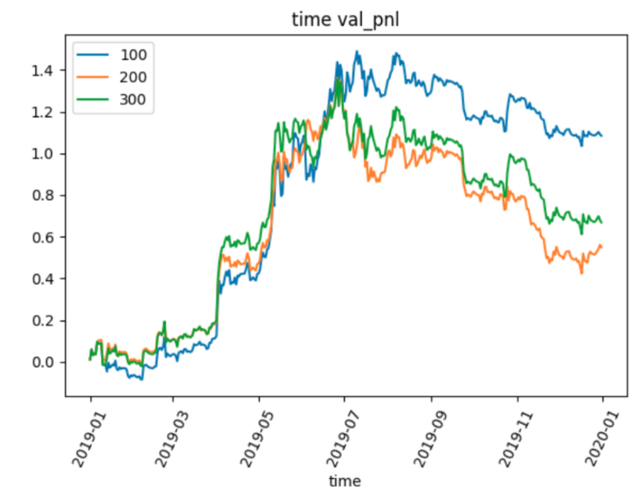
- Future could include training models end-to-end to directly predict the target label for trading.

Questions

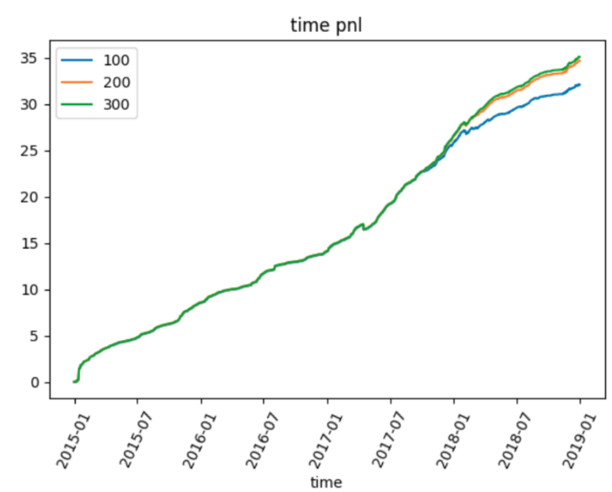
Results

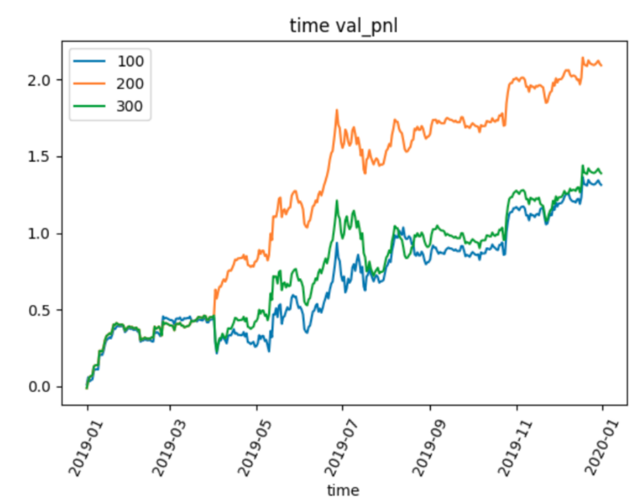
Price



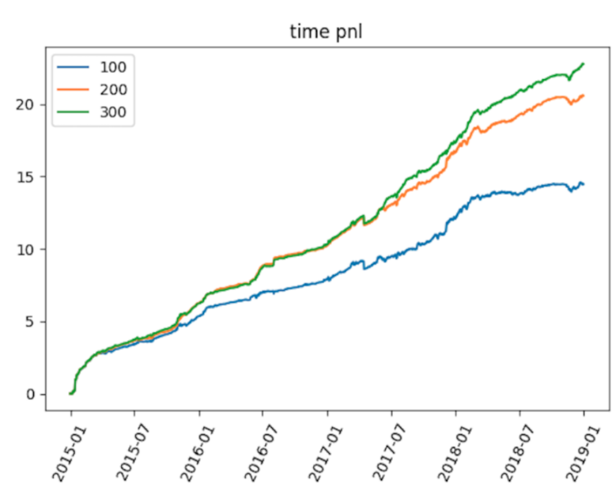


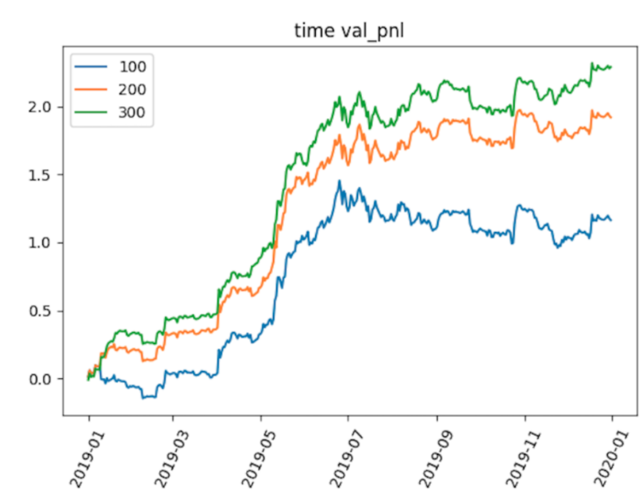
Sentiment





Price + Sentiment





Results

Input Modality	**MLP**	**CNN**	**LSTM**
price	201%	219%	214%
sentiment	221%	**228%**	222%
price & sentiment	**224%**	**228%**	**224%**