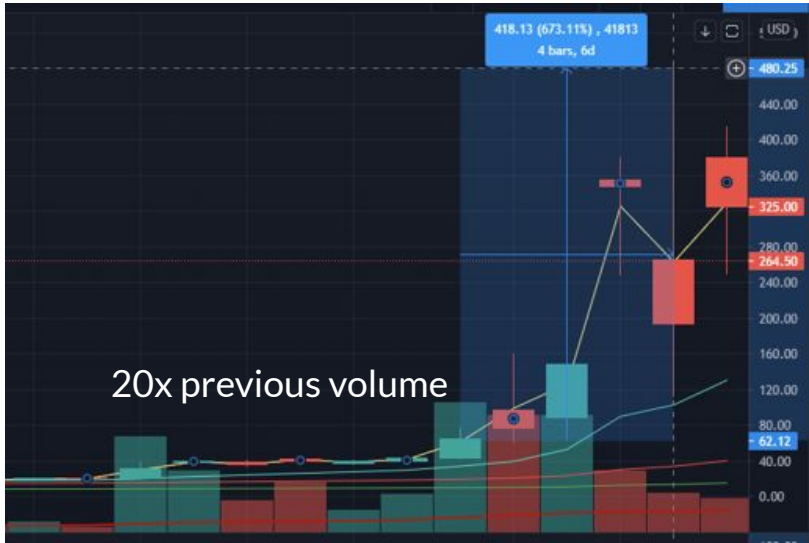
Finding high volume stocks and predicting price

Matthew Cheng| Bohdan Gilsevic

Is it too late to board the rocket?

- >600% over few days
- Extreme trading volume + short squeeze
- Find similar stocks automatically (1)...
- … predict if price will go up tomorrow (2).



Low volume

1. Stock Scanner

Scanning for high volume stocks:

➔ **Historic data**
Mean and standard deviation of volume in past 30 days

➔ **Now data**
Calculate z-score (number of standard deviation from mean)

➔ **Threshold**
If today's z-score is bigger than 5, flag it.

On a typical day for nasdaq, how many such stocks?

Tip:

NASDAQ: 3881

NYSE: 3044.

On a typical day for nasdaq, how many such stocks?

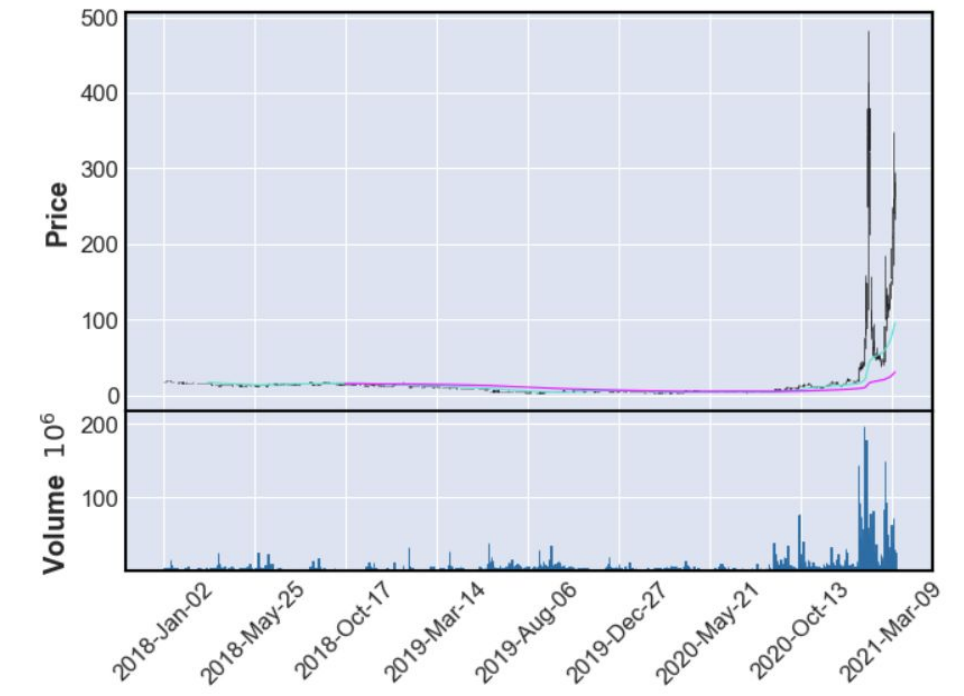
Just ~50 (1%)

Tip:

NASDAQ: 3881

NYSE: 3044.

High Volume
High Risk



2. Price Predictor

Will the price go up tomorrow?

➔ **Objective**

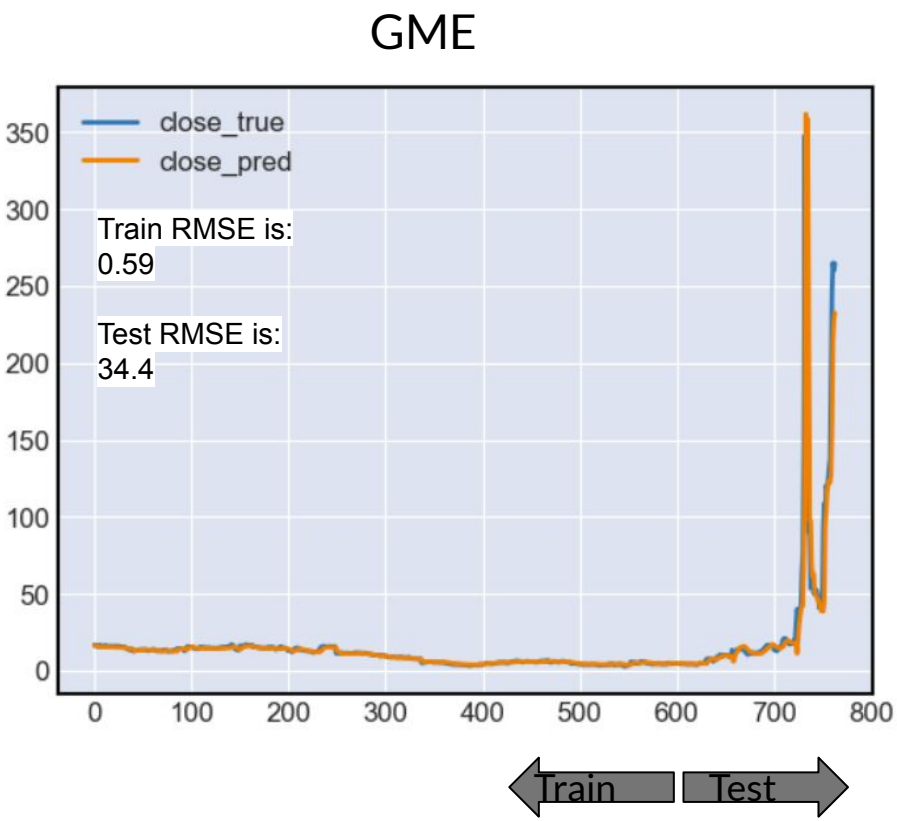
Do we predict the **price** or the **up/down** label?

➔ **Features**

What features to use?

➔ **Models**

What models to use for inference?



Case 1

Input: OHLCV for past 20 days

Output: C for tomorrow.

Model: LSTM + Linear

On average: predicted price deviates by ~$34 from true price.

Up/down label accuracy: 50%

– Case 2

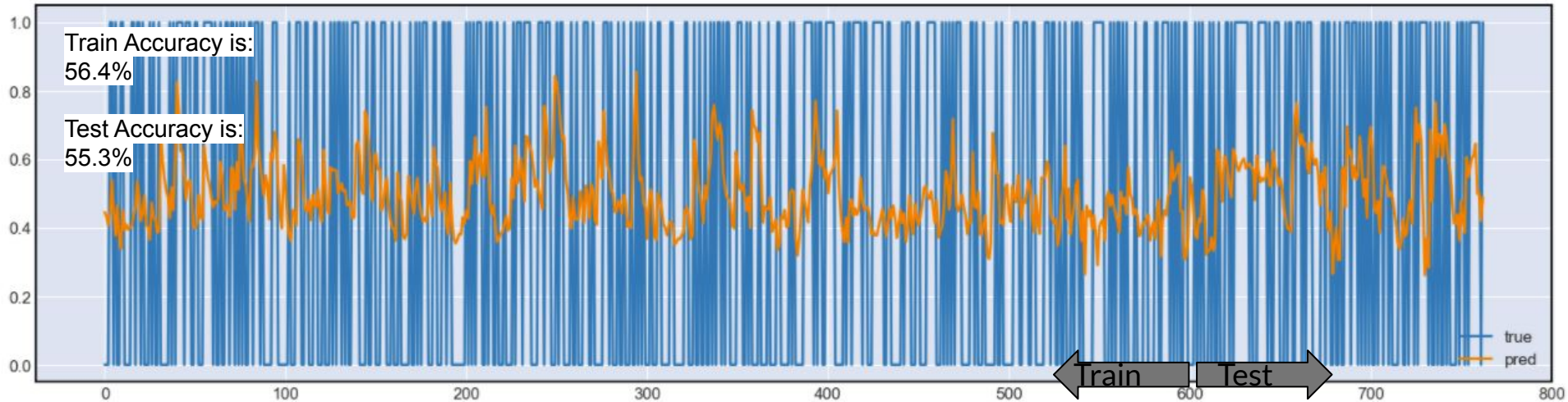
Input: OHLCV for past 20 days

Output: up/down for tomorrow.

Model: LSTM + Linear + sigmoid

On average: you could win 55%.

If you only bought on days y_pred = 1
Sell at next day close.
Win => 66.6%



Too basic...

3. Next?

➔ **More powerful features**

Volume driven by **hype**. Hype driven by people; retail / institutional. **Social media** data valuable. Indicate bullish/ bearish %

➔ **More powerful models**

Seq2Seq models, attention mechanism, transformer models, **HUGE** success in NLP (e.g. google translate). Could be leveraged in **price forecasting**