Stock Price Forecasting

Menghong Huang, Zhaosheng Li

Overview

- Background

- Proposal

- Methods

- Results

- Conclusion

- Future Work

Background

- **Relevant Literature**

    - [1] uses support vector regression (SVR) with windowing function

    - [2] adapts back propagation neural network (BPNN)

    - [3] implements long short–term memory (LSTM)

- **Data**

    - GameStop stock prices (GME) from Feb 2002 to Jan 2021.

    - Features: Open price, close price, high price, low price, volume, and adjusted close price

    - Target : close price
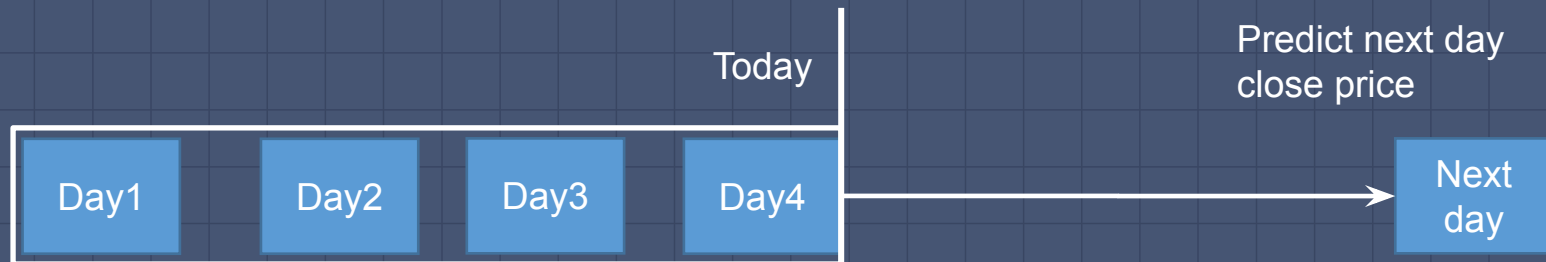
    - 80% for training set, 20% for test set
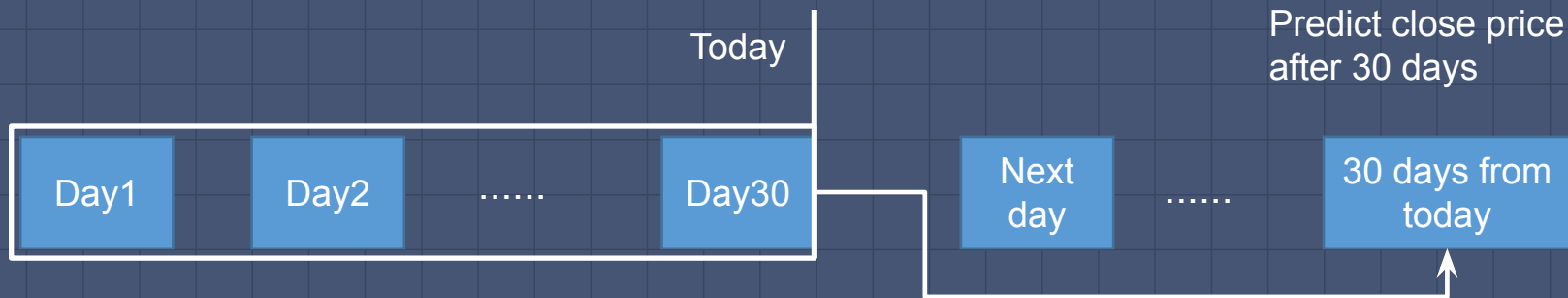
Proposal

- **Methods**

  - Linear Regression (LR)

  - Support Vector Regressor (SVR)

  - Backpropagation Neural Network (BPNN)

  - Basic Recurrent Neural Network (RNN)

  - Long Short-Term Memory (LSTM)

  - Gated Recurrent Units (GRU)

Proposal (cont.)

□ **Next-day Forecasting**

Today

| Day1 | Day2 | Day3 | Day4 |

Predict next day close price

Next day

□ **30-day Forecasting**

Today

Predict close price after 30 days

| Day1 | Day2 | ...... | Day30 |

Next day ...... 30 days from today

Proposal (cont.)

□ **Evaluation**

1. **RMSE:** The difference between actual price and predict price

2. **R^2:** coefficient of determination, the percentage of price variation can be explained by Inputs variation through regression models

3. **Accuracy of price changing direction:** correctly predict price going up or going down for each day
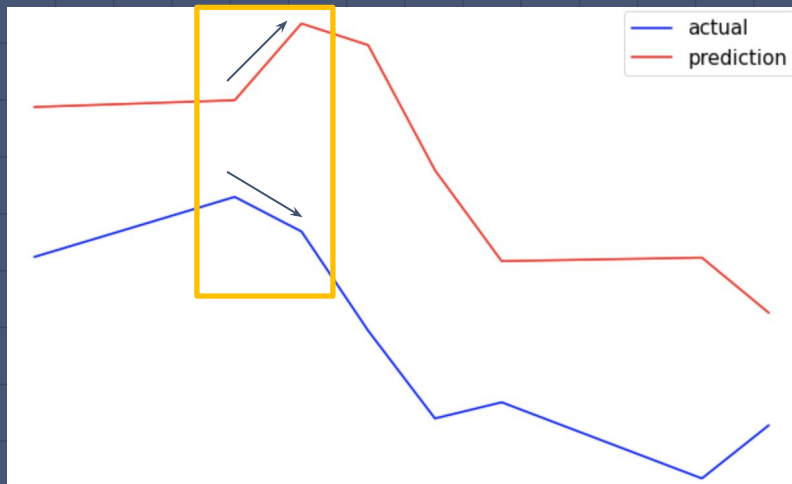
Methods

LR   SVR   BPNN

Basic RNN   LSTM   GRU

Methods -- LR

- Tuning parameters is unnecessary

- Determine the best fit model with intercept and slope

- Minimize the cost function

- Only accepts 1D input

- Input: Close price on today

[4]

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$* n$ is the number of data points
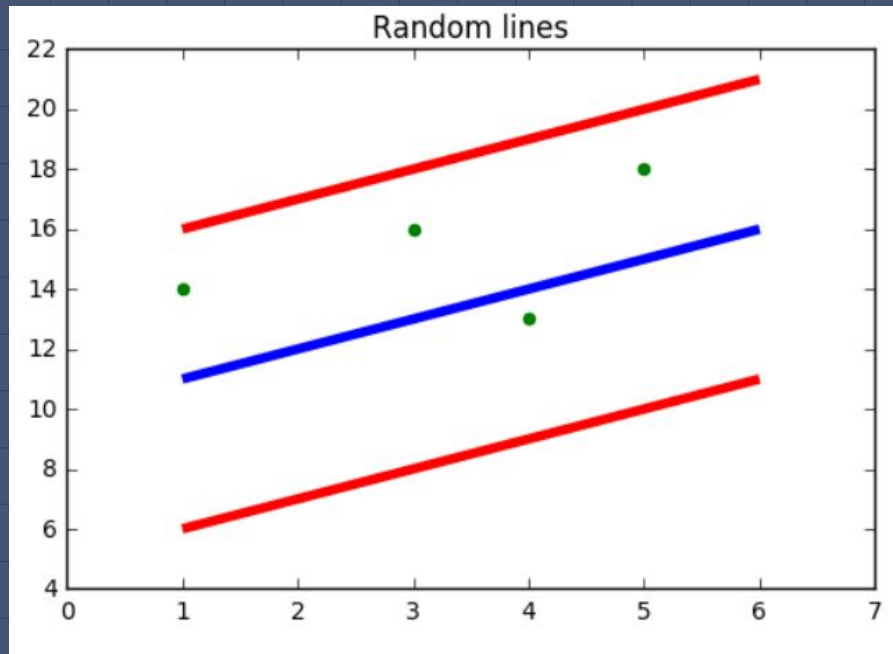$* Y_i$ represents observed values
$* \hat{Y}_i$ represents predicted values

Methods -- SVR

□ Combo of Support Vector Machine (SVM) and Regression

□ Fit the error within a certain threshold

[4]

□ Kernel: RBF (Radial Basis Function)

□ Tune penalty parameter (C) and γ

□ Input: Close price on today

[4]

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

Random lines

Methods -- BPNN

For Next_day Forecasting
- ▫ 3 fully-connected layers
- ▫ Input: Today's 6 feature + previous 3 days close price, scaled by standard scaler.
- ▫ Output: Next-day close price
- ▫ Activation: Relu
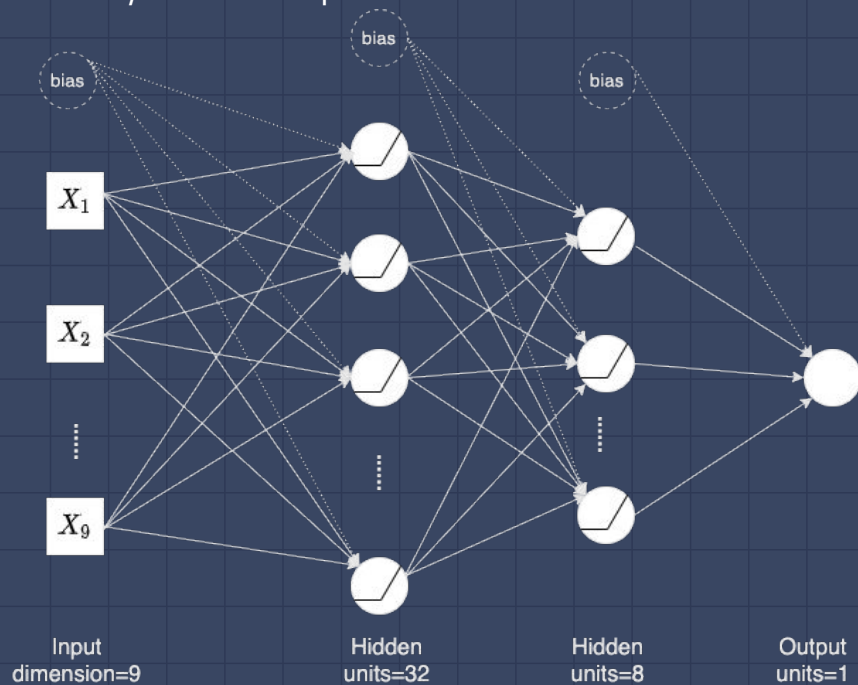- ▫ Loss: Mean Squared Error
- ▫ Optimizer: Adam

For 30_day Forecasting
- ▫ Input: previous 30-day close prices, standard scaled. Input dimension is 30.
- ▫ Output: 30-day afterwards close price

Why BPNN:
Besides regression models, we'd like to use deep learning tools.
BPNN is the fundamental one with good ability to identify non-linear pattern

BPNN Structure

Methods -- BPNN (cont.)

Special Input data setting:
- to make BPNN to learn the relationship of data in time sequence
- keep the low model complexity

Inputs of Next_day Forecasting

date	open_price	high_price	low_price	close_price	volume	adjclose_price	close_price_lag1 Yesterday	close_price_lag2 2 days before	close_price_lag3 3 days before
2021–01–07	18.47	19.45	18.02	18.08	6129300.00	18.08	18.36	17.37	17.25
2021–01–08	18.18	18.30	17.08	17.69	6464500.00	17.69	18.08	18.36	17.37
2021–01–11	19.41	20.65	19.01	19.94	14927600.00	19.94	17.69	18.08	18.36
2021–01–12	19.96	20.40	19.32	19.95	7060700.00	19.95	19.94	17.69	18.08

Inputs of 30_day Forecasting

date	Day1	Day2	Day3	Day4	Day5	Day6	...	Day24	Day25	Day26	Day27	Day28	Day29	Day30
					Close Price of 30 days									
2002–04–01	10.00	9.95	9.55	9.88	9.85	9.68	...	9.01	9.38	9.65	9.86	9.99	10.30	10.10

Methods -- Basic RNN

- A many-to-one RNN structure, 3 hidden RNN layers with both 64 units
- Next-day forecasting: Input 4-day sequential data with all 6 features, output next-day close price
- 30-day forecasting: look back 30-day data as input, output the next 30th close price
- Recurrent dropout 0.2 is taken to mitigate the gradient vanishing
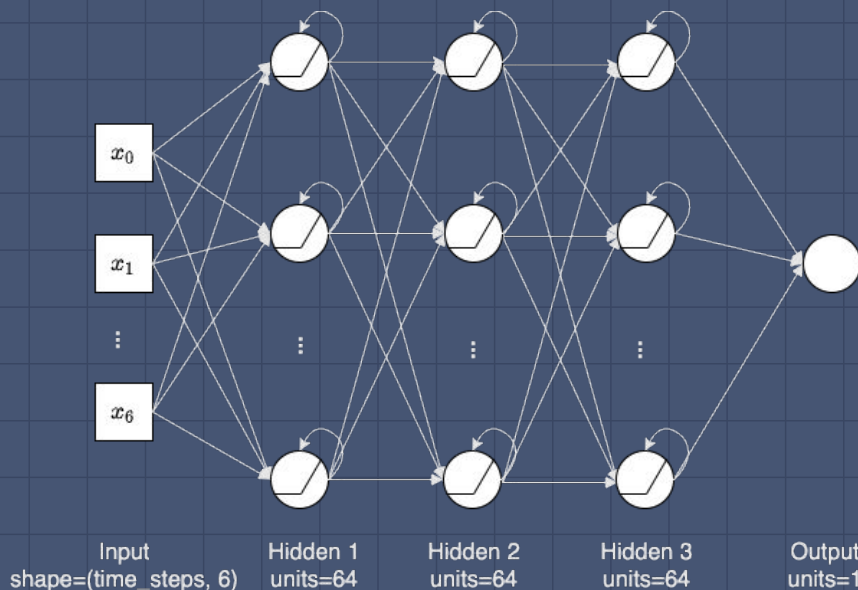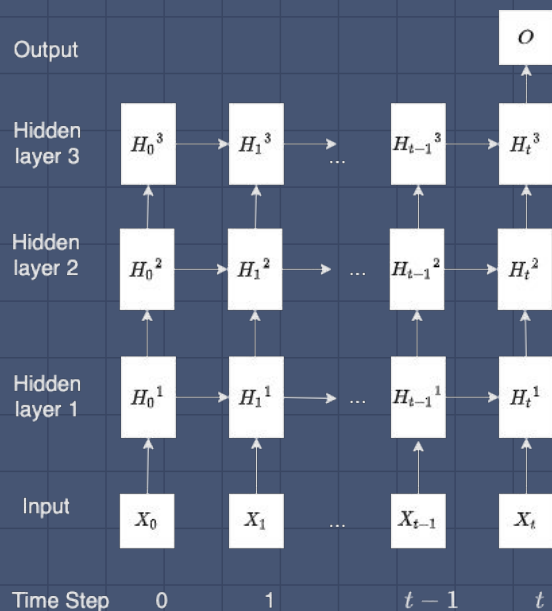
Methods -- LSTM & GRU

- Many-to-one architecture
- Avoid overfitting
  - Normalize data into the range [0, 1]
  - Dropout layer
- Input
  - Next-day forecast: latest 4-day data entries, including today
  - 30-day forecast: latest 30-day data entries, including today

Input

Units = 64

Dropout(0.2)

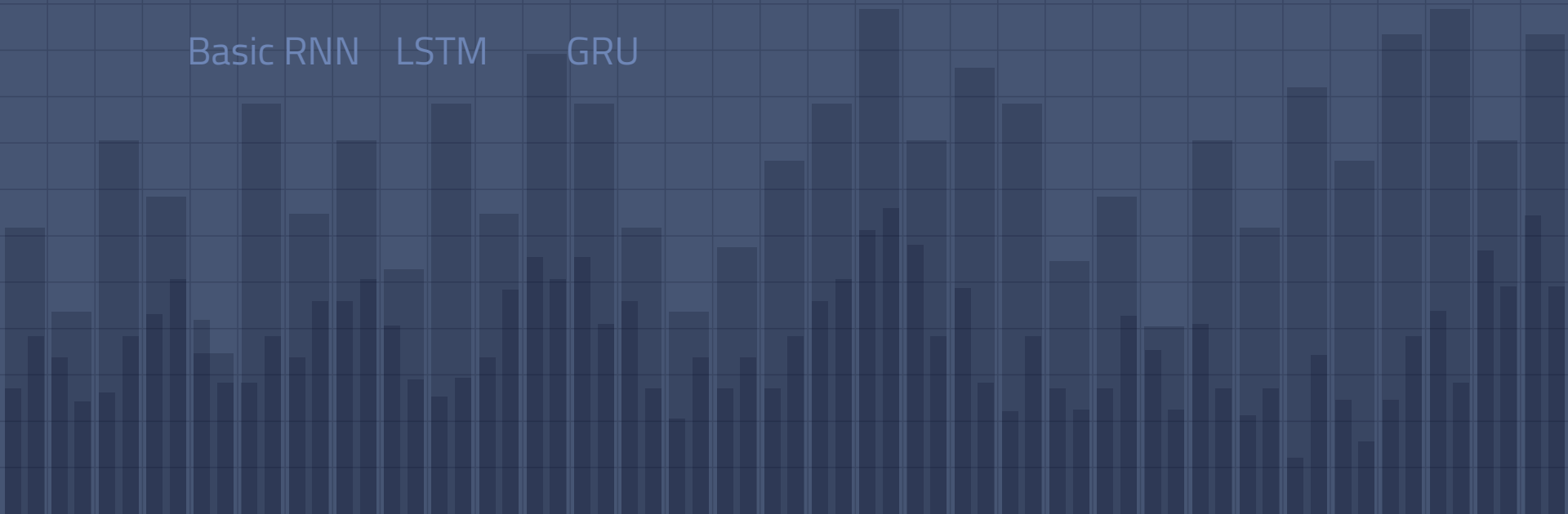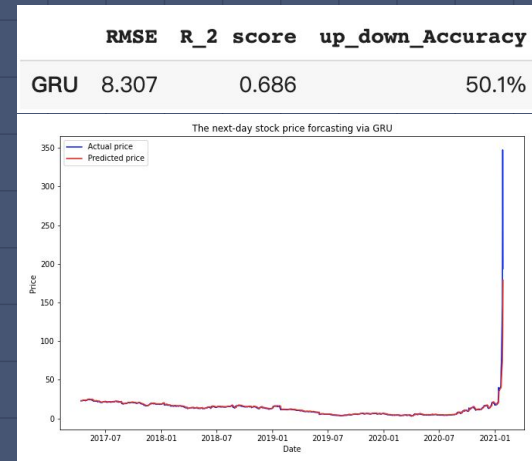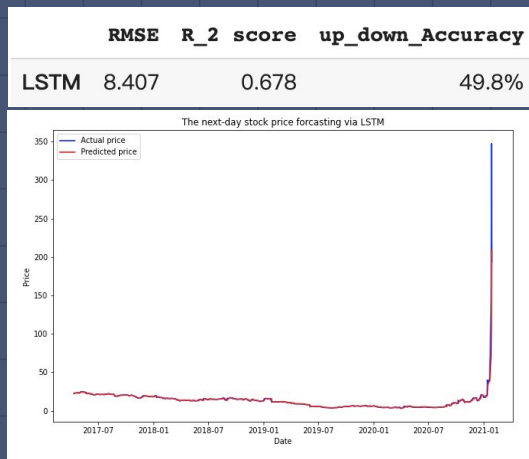Units = 64

Dropout(0.2)

Units = 64

Dropout(0.2)

Output (1 Neuron)
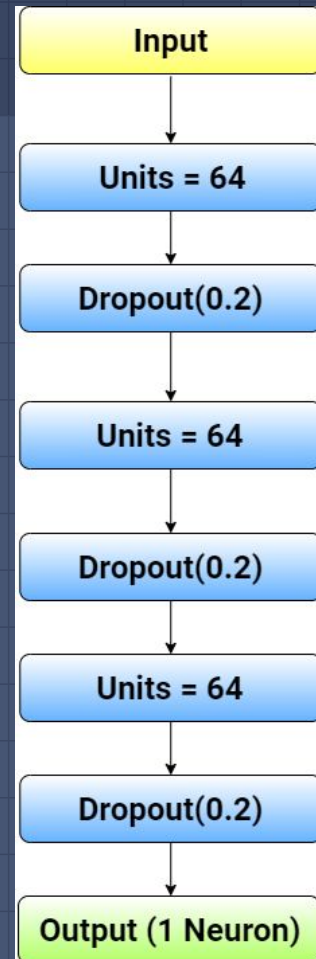
Results

LR   SVR   BPNN

Basic RNN   LSTM   GRU

Result -- Next-day Forecasting

	LR	SVR	BPNN	Basic RNN	LSTM	GRU
RMSE	8.537	12.01	7.485	7.705	8.407	8.307
R^2	0.668	0.344	0.745	0.730	0.678	0.686
Accuracy	50.1%	48.6%	50.0%	50.9%	49.8%	50.1%

Result -- Next-day Forecasting
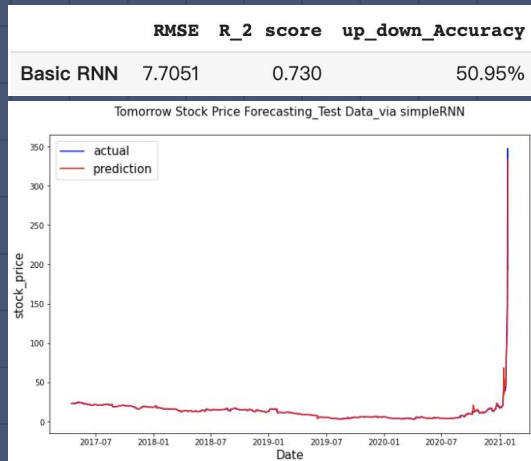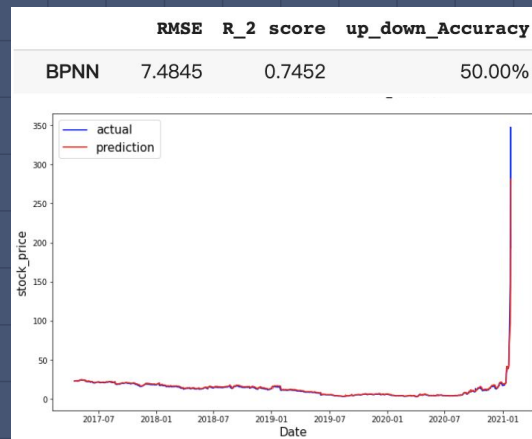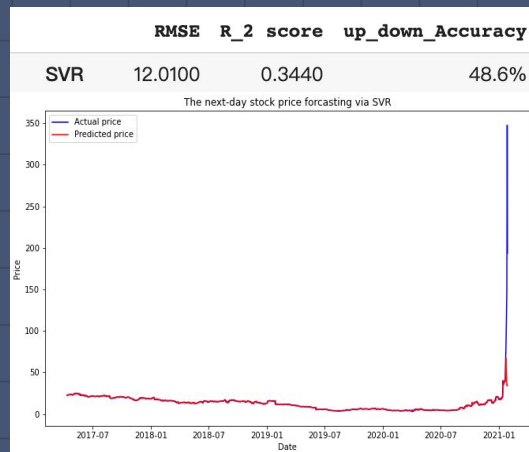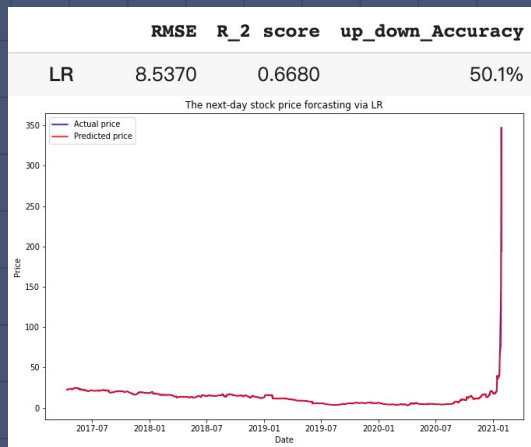
Result -- 30-day Forecasting

	LR	SVR	BPNN	Basic RNN	LSTM	GRU
RMSE	13.60	13.52	13.69	13.62	13.71	13.74
R^2	0.161	0.171	0.153	0.162	0.150	0.145
Accuracy	51.3%	48.3%	51.2%	49.4%	48.7%	50.3%

Result -- 30-day Forecasting

	RMSE	R_2 score	up_down_Accuracy
LR	13.60	0.1610	51.3%

The 30-day stock price forcasting via LR
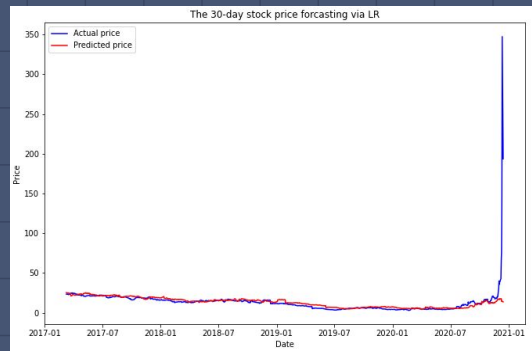
	RMSE	R_2 score	up_down_Accuracy
SVR	13.52	0.1710	48.3%

The 30-day stock price forcasting via SVR
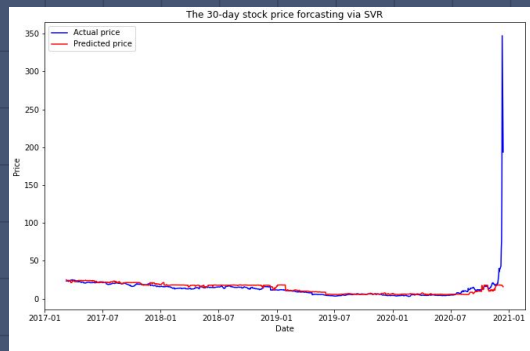
	RMSE	R_2 score	up_down_Accuracy
BPNN	13.69	0.1528	51.22%

Prediction VS Actual 30-Day-after Close Price_Test Data
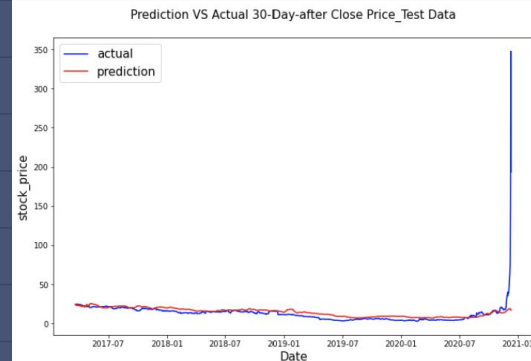
	RMSE	R_2 score	up_down_Accuracy
Basic RNN	13.62	0.1615	49.42%

The 30-Day_after Stock Price Forecasting_Test Data_via simpleRNN
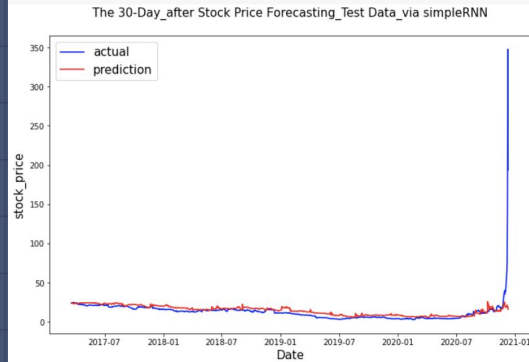
	RMSE	R_2 score	up_down_Accuracy
LSTM	13.71	0.150	48.7%

The 30-day stock price forcasting via LSTM
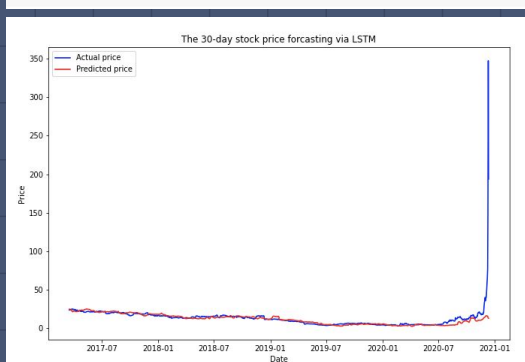
	RMSE	R_2 score	up_down_Accuracy
GRU	13.74	0.145	50.3%
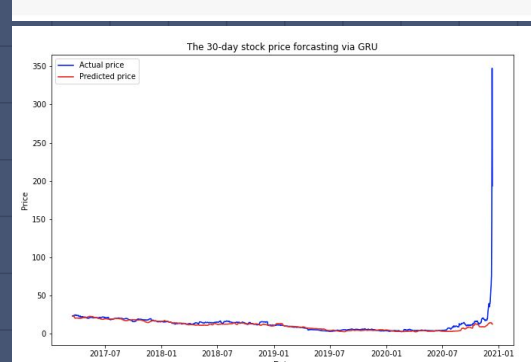
The 30-day stock price forcasting via GRU

But…

[5]

GameStop's stock is going through some stuff. You can thank Reddit

By Allison Morrow, CNN Business
Updated 7:56 PM ET, Mon January 25, 2021

[6]

GameStop frenzy explained: How small investors on Reddit took on Wall Street

By Thyagaraju Adinarayan · Reuters
Posted January 29, 2021 8:51 am · Updated January 29, 2021 8:55 am

GME Close Price

GME Close Price (cont.)

GameStop close price in the last 20 days
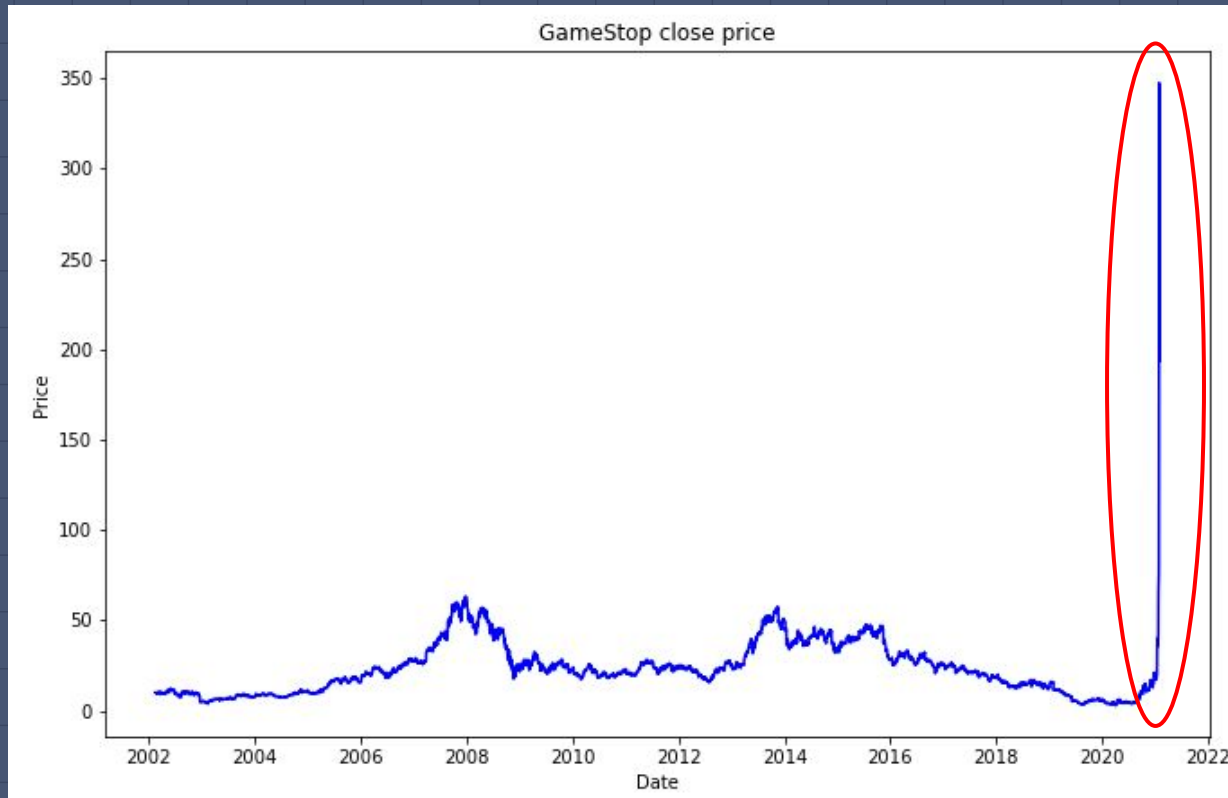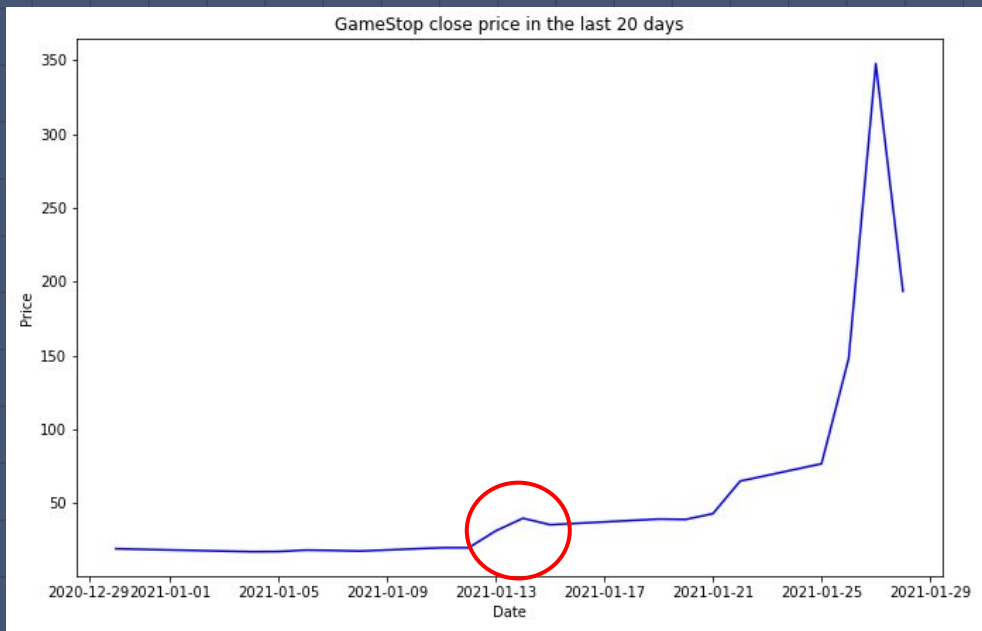
Date	Price
2021-01-04	17.25
2021-01-05	17.37
2021-01-06	18.36
2021-01-07	18.08
2021-01-08	17.69
2021-01-11	19.94
2021-01-12	19.95
2021-01-13	31.4
2021-01-14	39.91
2021-01-15	35.5
2021-01-19	39.36
2021-01-20	39.12
2021-01-21	43.03
2021-01-22	65.01
2021-01-25	76.79
2021-01-26	147.98
2021-01-27	347.51
2021-01-28	193.6

Result -- Next-day Forecasting

	LR	SVR	BPNN	Basic RNN	LSTM	GRU
RMSE	0.502	0.523	0.819	0.669	0.583	0.663
R^2	0.993	0.993	0.982	0.988	0.991	0.988
Accuracy	50.1%	48.6%	49.9%	50.9%	49.6%	49.8%

Result -- Next-day Forecasting

	RMSE	R_2_score	up_down_Accuracy
LR	0.5018	0.9933	50.10%

The next-day stock price forcasting without the last 11 days via LR
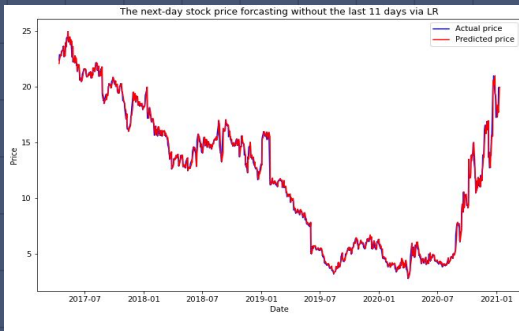
	RMSE	R_2_score	up_down_Accuracy
SVR	0.5275	0.9926	48.64%

The next-day stock price forcasting without the last 11 days via SVR

	RMSE	R_2_score	up_down_Accuracy
BPNN	0.8191	0.9820	49.92%

Prediction VS Actual Tomorrow Close Price_Test Data_without last 11 data

	RMSE	R_2_score	up_down_Accuracy
Basic RNN	0.6689	0.9880	50.90%

Tomorrow Stock Price Forecasting_Test Data_without last 11 data_via simpleRNN
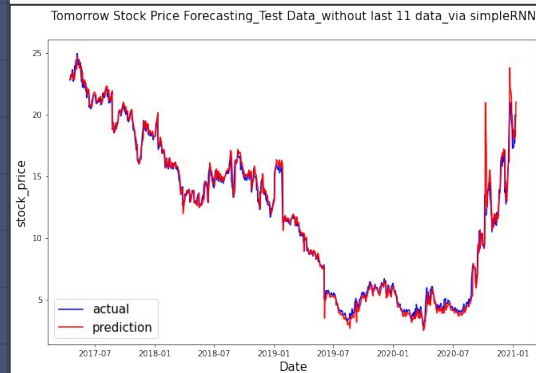
	RMSE	R_2_score	up_down_Accuracy
LSTM	0.5827	0.9909	49.58%

The next-day stock price forcasting without the last 11 days via LSTM
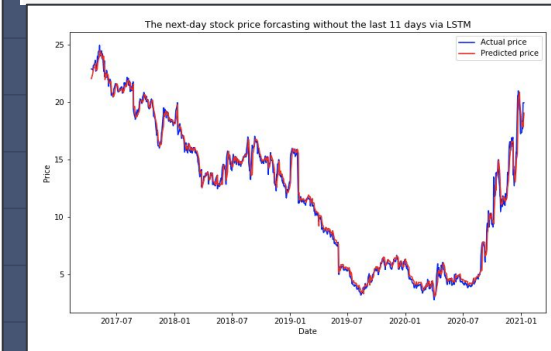
	RMSE	R_2_score	up_down_Accuracy
GRU	0.6628	0.9882	49.79%

The next-day stock price forcasting without the last 11 days via GRU
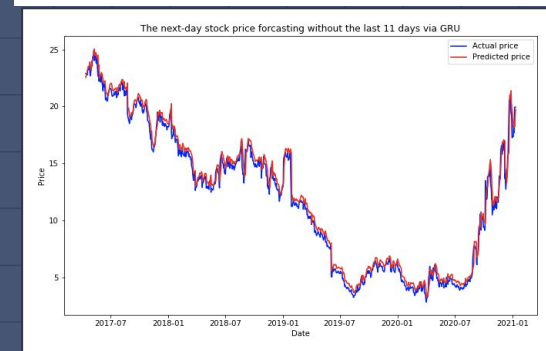
Result -- 30-day Forecasting

	LR	SVR	BPNN	Basic RNN	LSTM	GRU
RMSE	2.365	2.809	3.539	3.509	2.407	2.438
R²	0.848	0.786	0.654	0.660	0.841	0.836
Accuracy	51.3%	48.3%	51.1%	49.1%	48.9%	50.3%

Result -- 30-day Forecasting

	RMSE	R_2 score	up_down_Accuracy
LR	2.3650	0.8480	51.3%

The next-day stock price forcasting without the last 11 days via LR

	RMSE	R_2 score	up_down_Accuracy
SVR	2.809	0.786	48.3%

The next-day stock price forcasting without the last 11 days via SVR

	RMSE	R_2 score	up_down_Accuracy
BPNN	3.5390	0.6540	51.22%

Prediction VS Actual 30-Day-after Close Price_Test Data

	RMSE	R_2 score	up_down_Accuracy
Basic RNN	3.5091	0.6599	49.14%

The 30-Day_after Stock Price Forecasting_Test Data_without last 11 data_via simpleRNN

	RMSE	R_2 score	up_down_Accuracy
LSTM	2.407	0.841	48.9%

The 30-day stock price forcasting without the last 11 days via LSTM

	RMSE	R_2 score	up_down_Accuracy
GRU	2.438	0.836	50.3%

The 30-day stock price forcasting without the last 11 days via GRU

Conclusion

- The close prices on the adjacent two days have strong linear relationship

- SVR has looser tolerance for error, compared to simple LR

- BPNN have shortage in long-term prediction. Without most close dependencies, the prediction on next 30 days is weaken.

- Basic RNN can well capture the price pattern. However, it has gradient vanishing problem in 30-day forecasting since 30 day is comparatively long time sequence.

- A spike on the stock price may highly degrade the model performance

Future Work

- Implement more various machine learning models

- Conduct mid-term or long-term forecast

- Research on how to improve the trend accuracy

Thank you

Any questions?

References

- [1] P. Meesad and R. I. Rasel, "Predicting stock market price using support vector regression," 2013 International Conference on Informatics, Electronics and Vision (ICIEV), pp. 1-6, 2013.

- [2] W. Ma, Y. Wang, and N. Dong, "Study on stock price prediction based on bp neural network," 2010 IEEE International Conference on Emergency Management and Management Sciences, 2010.

- [3] J. Qiu, B. Wang, and C. Zhou, "Forecasting stock prices with long-short term memory neural network based on attention mechanism," PLoS ONE, vol. 15, 2020.

- [4] D. Scatterday, "Walking through Support Vector Regression and LSTMs with stock price prediction," *Medium*, 18-Sep-2019. [Online]. Available: https://towardsdatascience.com/walking-through-support-vector-regression-and-lstms-with-stock-price-prediction-45e11b620650 . [Accessed: 19-Apr-2021].

References (cont.)

▫ [5] A. Morrow, "GameStop's stock is going through some stuff. You can thank Reddit," *CNN Business*, 25-Jan-2021.

▫ [6] T. Adinarayan, "GameStop frenzy explained: How small investors on Reddit took on Wall Street," *Global News*, 29-Jan-2021.

▫ [7] 齊克用, "全球股市潛伏危機？齊克用：中期上漲趨勢仍未結束 - 今周刊," *專業財經雜誌新聞網*, 07-Mar-2019. [Online]. Available: https://www.businesstoday.com.tw/article/category/80402/post/201903070027/%E5%85%A8%E7%90%83%E8%82%A1%E5%B8%82%E6%BD%9B%E4%BC%8F%E5%8D%B1%E6%A9%9F%EF%BC%9F%E9%BD%8A%E5%85%8B%E7%94%A8%EF%BC%9A%E4%B8%AD%E6%9C%9F%E4%B8%8A%E6%BC%B2%E8%B6%A8%E5%8B%A2%E4%BB%8D%E6%9C%AA%E7%B5%90%E6%9D%9F . [Accessed: 20-Apr-2021].