PROJECT REPORT

Stock Movement Prediction

Language and Image Processing
Rudayasekaran Panneer Selvam
B87762

INTRODUCTION

Predicting future stock price movement based on three sources of data.

This project presents a methodology to use the content of news information, stock chart's images and actual stock values to predict stock price performance in future.

The main network is a combination of recurrent , convolutional and feed forward networks.

DATASET

	date	open	high	low	close	volume
0	2006-12-01	13.114285	13.190000	12.871428	13.045714	198769900
1	2006-12-04	13.125714	13.150000	12.928572	13.017143	177384200
2	2006-12-05	13.092857	13.190000	12.981428	13.038571	165709600
3	2006-12-06	12.948571	13.055715	12.810000	12.832857	159546100
4	2006-12-07	12.861428	12.928572	12.414286	12.434286	251206900

Stock Price Data

Extracted from MT5 trading terminal.

2006 to 2019, hourly intervals.

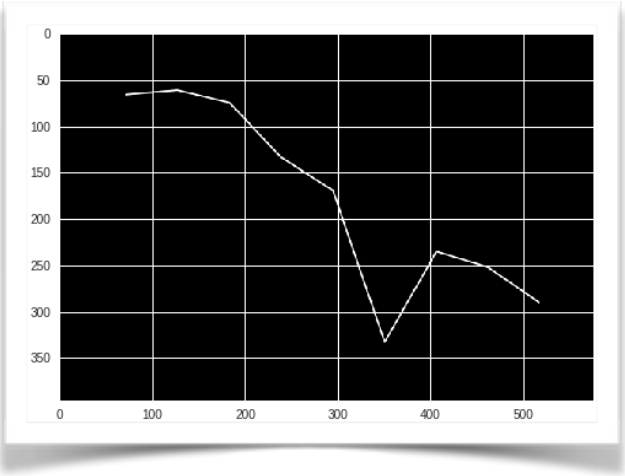


Chart Images Data

Generated using above stock values.

Shape - 396, 576, 1

In example; today's chart illustrates the movements of previous 10 days.

"Sales of iPods and iTunes Not Much in Sync : The numbers suggest that iPods are not driving iTunes sales as much as early supporters may have expected.Name That Source : Decades on, a Vietnam War-era mystery has been revealed.In City's Trans Fat Ban, a Challenge Fit for a Chef : Many in the restaurant industry say they fear that they will not be able to replicate dishes that now exceed the limit on trans fats.After Nasdaq Chief's Irish Bash, a Long Hangover : Even as the Nasdaq Stock Market battles to acquire a recalcitrant London Stock Exchange, Nasdaq's chief executive is in a trans-Atlantic feud of his own over a week of medieval revelry in Ireland. According to the New York Daily News, Robert Greif...Lights! Camera! Advertising! : Squeezing Money From the Music : Major labels have begun demanding a cut of concert earnings or T-shirt, ring tone and merchandise revenue from new artists seeking record contracts.

News Data

Downloaded from Kaggle

Not labelled

Max length - around 4000

Average length - 1400

DATA PREPROCESS

News

Generated Polarity values using NLTK Sentiment Analyser.

Tokenisation performed and StopWords removed.

Padded shorter sentences to the size of maximum sentence length.

Glove embedding was used for word2idx and for embedding matrix.

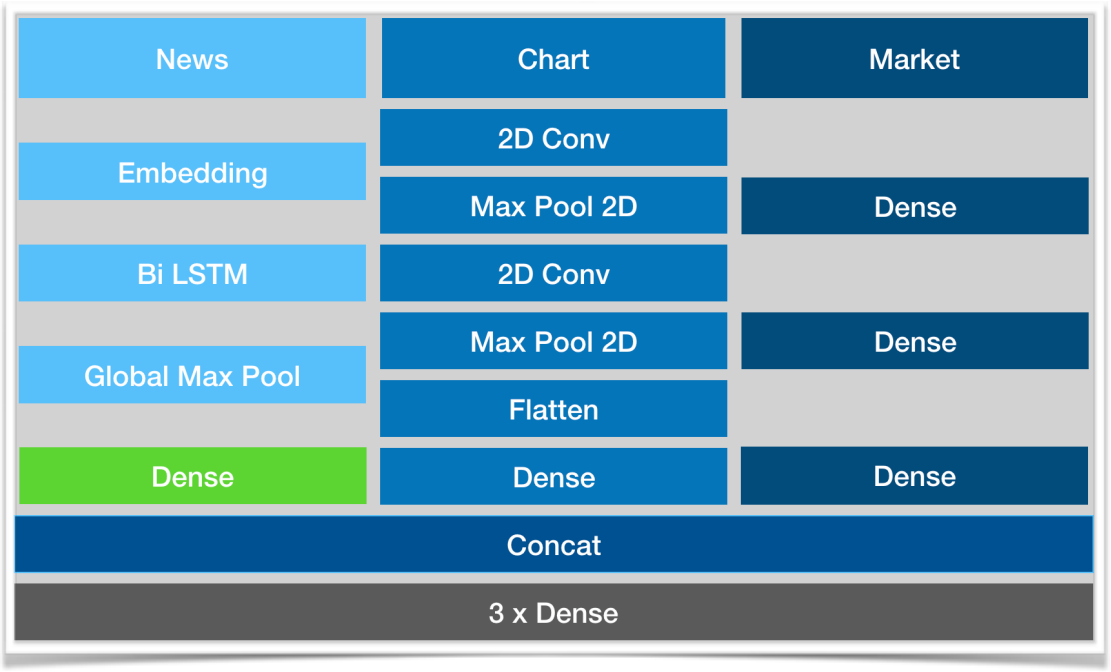
Images

Grayscaled all the images and reshaped.

Stock Values

Normalised all the numeric values using StandardScaler.

IMPLEMENTATION



Input 1: News Data

Input 2: Chart Images

Input 3: Numeric Stock Values

Output 1: Actual Prediction of the stock movement

Output 2: Dense Layer Output of News

My idea was to target the polarity values I generated for news data. So I created another output specifically for news and made it to target the polarity values, the dense layer which showed in green was the 2nd output and I still used it in the model in later layers to generate the actual prediction too. Basically useful for training the network, no use when do testing or go live. In other words, I just tried to do supervised training for news. And at the same time I wanted to do completely unsupervised approach for chart images, even though I have used very basic types of chart images.

Other Informations

MAX_SEQUENCE_LENGTH = 4089

MAX_VOCAB_SIZE = 40000

EMBEDDING_DIM = 200

num_words = 40000

batch_size=64

epochs=40

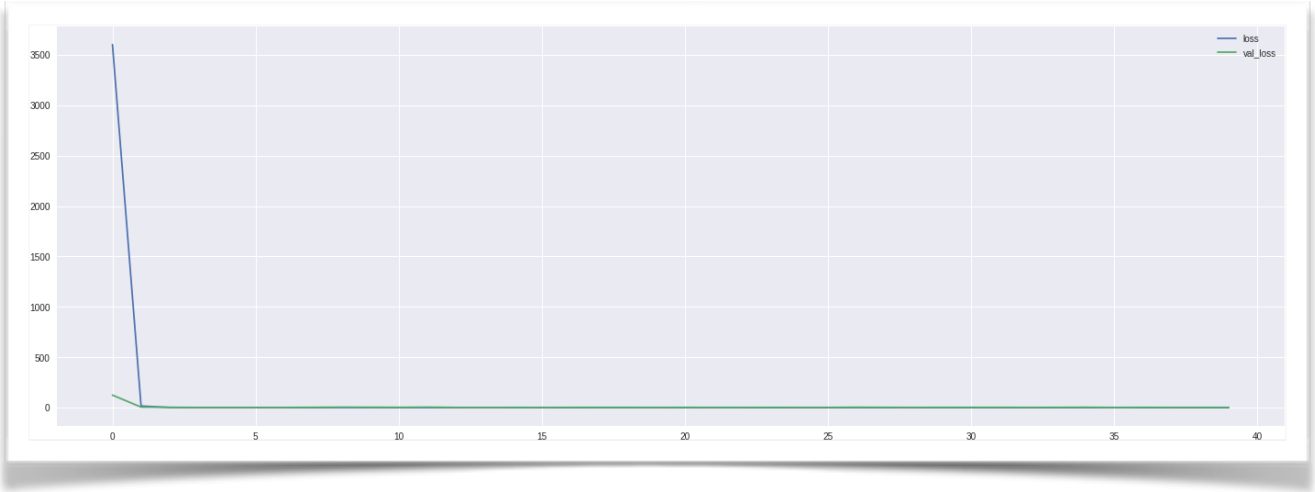
Optimiser - Adam

Loss - mean absolute error

I have used mean absolute error as the loss function, which will be useful to measure the difference between the actual stock value and the predicted stock value. In this scenario I got around 2.5 as the value for mae, which means the difference between actual and predicted is around 2.5 US dollars.

In trading: normally trader's success is measured by how many trades he/she has won in percentage. But this system does not predict any trading decisions, all it does is predict the movements. So in my opinion MAE is a reasonable loss function to benchmark the system.

TRAINING




```
Train on 2025 samples, validate on 226 samples
Epoch 1/40
2025/2025 [==============================] - 31s 15ms/step - loss: 3021.4496 - dense_24_loss: 3021.0232 - dense_17_loss: 0.4262 -
val_loss: 122.7476 - val_dense_24_loss: 122.3371 - val_dense_17_loss: 0.4105
Epoch 2/40
2025/2025 [==============================] - 29s 15ms/step - loss: 17.4626 - dense_24_loss: 17.0279 - dense_17_loss: 0.4347 -
val_loss: 10.5516 - val_dense_24_loss: 10.1173 - val_dense_17_loss: 0.4343
Epoch 3/40
2025/2025 [==============================] - 29s 15ms/step - loss: 3.1774 - dense_24_loss: 2.7881 - dense_17_loss: 0.3892 -
val_loss: 4.4081 - val_dense_24_loss: 4.0246 - val_dense_17_loss: 0.3835
Epoch 4/40
2025/2025 [==============================] - 29s 14ms/step - loss: 2.2624 - dense_24_loss: 1.8954 - dense_17_loss: 0.3669 -
val_loss: 3.3643 - val_dense_24_loss: 2.9866 - val_dense_17_loss: 0.3777
Epoch 5/40
2025/2025 [==============================] - 29s 14ms/step - loss: 2.0122 - dense_24_loss: 1.6696 - dense_17_loss: 0.3426 -
val_loss: 4.9686 - val_dense_24_loss: 4.5997 - val_dense_17_loss: 0.3688
Epoch 6/40
2025/2025 [==============================] - 29s 14ms/step - loss: 2.0782 - dense_24_loss: 1.7498 - dense_17_loss: 0.3284 -
val_loss: 4.1643 - val_dense_24_loss: 3.8004 - val_dense_17_loss: 0.3639
Epoch 7/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.9709 - dense_24_loss: 1.6543 - dense_17_loss: 0.3166 -
val_loss: 6.4598 - val_dense_24_loss: 6.0936 - val_dense_17_loss: 0.3661
Epoch 8/40
2025/2025 [==============================] - 29s 14ms/step - loss: 2.1209 - dense_24_loss: 1.8103 - dense_17_loss: 0.3106 -
val_loss: 3.9256 - val_dense_24_loss: 3.5572 - val_dense_17_loss: 0.3684
Epoch 9/40
2025/2025 [==============================] - 29s 15ms/step - loss: 2.0668 - dense_24_loss: 1.7632 - dense_17_loss: 0.3036 -
val_loss: 4.0471 - val_dense_24_loss: 3.6821 - val_dense_17_loss: 0.3650
Epoch 10/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.8998 - dense_24_loss: 1.6009 - dense_17_loss: 0.2989 -
val_loss: 3.5200 - val_dense_24_loss: 3.1462 - val_dense_17_loss: 0.3738
Epoch 11/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.8368 - dense_24_loss: 1.5402 - dense_17_loss: 0.2966 -
val_loss: 3.6502 - val_dense_24_loss: 3.2853 - val_dense_17_loss: 0.3649
Epoch 12/40
2025/2025 [==============================] - 29s 15ms/step - loss: 2.0175 - dense_24_loss: 1.7209 - dense_17_loss: 0.2966 -
val_loss: 3.5814 - val_dense_24_loss: 3.2136 - val_dense_17_loss: 0.3677
Epoch 13/40
2025/2025 [==============================] - 29s 15ms/step - loss: 2.1676 - dense_24_loss: 1.8638 - dense_17_loss: 0.3038 -
val_loss: 3.4034 - val_dense_24_loss: 3.0202 - val_dense_17_loss: 0.3832
Epoch 14/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.8718 - dense_24_loss: 1.5681 - dense_17_loss: 0.3037 -
val_loss: 3.6809 - val_dense_24_loss: 3.3065 - val_dense_17_loss: 0.3744
Epoch 15/40
2025/2025 [==============================] - 29s 14ms/step - loss: 2.0840 - dense_24_loss: 1.7761 - dense_17_loss: 0.3079 -
val_loss: 3.4611 - val_dense_24_loss: 3.0761 - val_dense_17_loss: 0.3850
Epoch 16/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.9100 - dense_24_loss: 1.6033 - dense_17_loss: 0.3068 -
val_loss: 4.2323 - val_dense_24_loss: 3.8394 - val_dense_17_loss: 0.3929
Epoch 17/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.9932 - dense_24_loss: 1.6767 - dense_17_loss: 0.3164 -
val_loss: 4.0639 - val_dense_24_loss: 3.6833 - val_dense_17_loss: 0.3805
Epoch 18/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.9185 - dense_24_loss: 1.5918 - dense_17_loss: 0.3268 -
val_loss: 4.9535 - val_dense_24_loss: 4.5559 - val_dense_17_loss: 0.3976
```

```
Epoch 19/40
2025/2025 [==============================] - 29s 15ms/step - loss: 2.0625 - dense_24_loss: 1.7356 - dense_17_loss: 0.3270 -
val_loss: 3.4485 - val_dense_24_loss: 3.0567 - val_dense_17_loss: 0.3917
Epoch 20/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.9108 - dense_24_loss: 1.5751 - dense_17_loss: 0.3357 -
val_loss: 4.6005 - val_dense_24_loss: 4.1949 - val_dense_17_loss: 0.4057
Epoch 21/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.9494 - dense_24_loss: 1.6074 - dense_17_loss: 0.3421 -
val_loss: 3.6987 - val_dense_24_loss: 3.3057 - val_dense_17_loss: 0.3931
Epoch 22/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.9324 - dense_24_loss: 1.5790 - dense_17_loss: 0.3534 -
val_loss: 3.7363 - val_dense_24_loss: 3.3366 - val_dense_17_loss: 0.3997
Epoch 23/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.8506 - dense_24_loss: 1.4872 - dense_17_loss: 0.3634 -
val_loss: 5.2569 - val_dense_24_loss: 4.8573 - val_dense_17_loss: 0.3996
Epoch 24/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.8945 - dense_24_loss: 1.5201 - dense_17_loss: 0.3744 -
val_loss: 3.5217 - val_dense_24_loss: 3.1411 - val_dense_17_loss: 0.3807
Epoch 25/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.8377 - dense_24_loss: 1.4564 - dense_17_loss: 0.3814 -
val_loss: 3.5241 - val_dense_24_loss: 3.1351 - val_dense_17_loss: 0.3890
Epoch 26/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.9568 - dense_24_loss: 1.5678 - dense_17_loss: 0.3890 -
val_loss: 7.0759 - val_dense_24_loss: 6.6845 - val_dense_17_loss: 0.3915
Epoch 27/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.8091 - dense_24_loss: 1.4072 - dense_17_loss: 0.4018 -
val_loss: 3.5203 - val_dense_24_loss: 3.0967 - val_dense_17_loss: 0.4236
Epoch 28/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.6878 - dense_24_loss: 1.2691 - dense_17_loss: 0.4187 -
val_loss: 3.5527 - val_dense_24_loss: 3.1338 - val_dense_17_loss: 0.4188
Epoch 29/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.6593 - dense_24_loss: 1.2128 - dense_17_loss: 0.4465 -
val_loss: 3.6530 - val_dense_24_loss: 3.2280 - val_dense_17_loss: 0.4250
Epoch 30/40
2025/2025 [==============================] - 30s 15ms/step - loss: 2.0197 - dense_24_loss: 1.5679 - dense_17_loss: 0.4518 -
val_loss: 5.8136 - val_dense_24_loss: 5.4107 - val_dense_17_loss: 0.4029
Epoch 31/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.8134 - dense_24_loss: 1.3671 - dense_17_loss: 0.4462 -
val_loss: 3.9126 - val_dense_24_loss: 3.4945 - val_dense_17_loss: 0.4181
Epoch 32/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.6881 - dense_24_loss: 1.2370 - dense_17_loss: 0.4511 -
val_loss: 3.6388 - val_dense_24_loss: 3.1927 - val_dense_17_loss: 0.4461
Epoch 33/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.5977 - dense_24_loss: 1.1154 - dense_17_loss: 0.4823 -
val_loss: 3.5926 - val_dense_24_loss: 3.1668 - val_dense_17_loss: 0.4258
Epoch 34/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.7652 - dense_24_loss: 1.2878 - dense_17_loss: 0.4774 -
val_loss: 4.3798 - val_dense_24_loss: 3.9537 - val_dense_17_loss: 0.4261
Epoch 35/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.5407 - dense_24_loss: 1.0614 - dense_17_loss: 0.4793 -
val_loss: 4.6963 - val_dense_24_loss: 4.2553 - val_dense_17_loss: 0.4410
Epoch 36/40
2025/2025 [==============================] - 29s 14ms/step - loss: 1.8937 - dense_24_loss: 1.4109 - dense_17_loss: 0.4829 -
val_loss: 4.3524 - val_dense_24_loss: 3.9242 - val_dense_17_loss: 0.4281
Epoch 37/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.6675 - dense_24_loss: 1.1871 - dense_17_loss: 0.4804 -
val_loss: 3.5439 - val_dense_24_loss: 3.0999 - val_dense_17_loss: 0.4440
Epoch 38/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.6268 - dense_24_loss: 1.1334 - dense_17_loss: 0.4934 -
val_loss: 3.8259 - val_dense_24_loss: 3.3914 - val_dense_17_loss: 0.4345
Epoch 39/40
2025/2025 [==============================] - 30s 15ms/step - loss: 1.6156 - dense_24_loss: 1.1269 - dense_17_loss: 0.4887 -
val_loss: 3.5079 - val_dense_24_loss: 3.0625 - val_dense_17_loss: 0.4454
Epoch 40/40
2025/2025 [==============================] - 29s 15ms/step - loss: 1.6429 - dense_24_loss: 1.1543 - dense_17_loss: 0.4886 -
val_loss: 3.5303 - val_dense_24_loss: 3.0844 - val_dense_17_loss: 0.4459
```

PREDICTIONS & CONCLUSION

mean_absolute_error - 2.5613863826568726

This network can be improved by using:

- Labelled news data
- Domain related news data
- Chart images with more complex information, such as trend, supply and demand zones.
- Custom layer with Arima capabilities.