Final Project Results

Diego Ferrer | Owen Farthing

Overview

Recap

- Finding correlations between Robinhood user data and stock prices
- Using correlated stocks to model and make predictions
- Analyzing results to determine the usefulness of Robinhood data

Changes

- Added K-Nearest Neighbor for modeling and prediction

Data Pipeline

Preprocessing

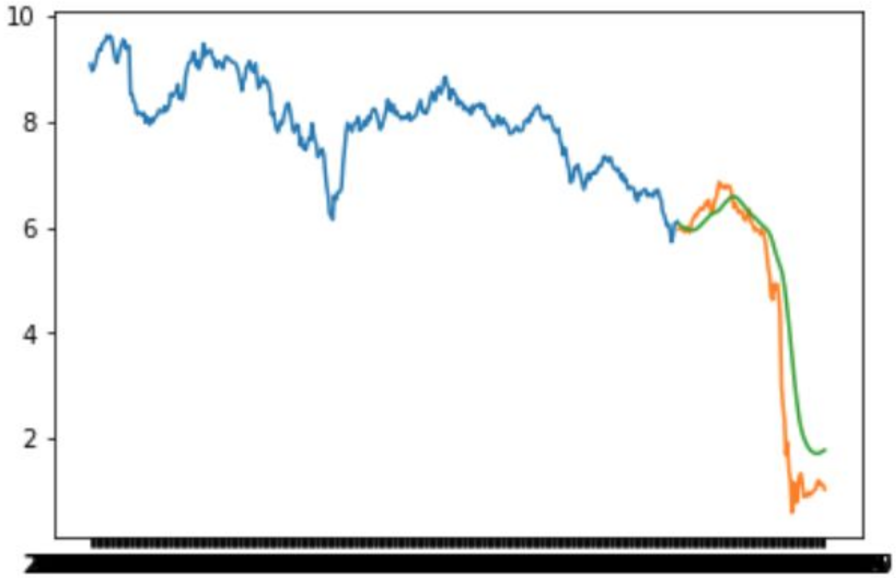
- Loaded files into memory
- Concatenated and formatted them
- Calculated averages and correlation
- Applied correlation threshold to filter stocks

Analysis

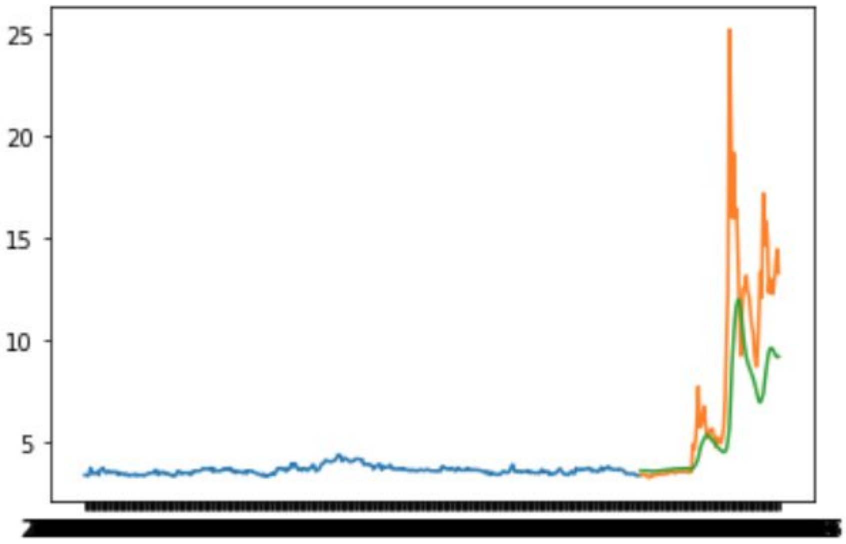
- Trained and tested various models on combined data
- Chose features useful for specific models
- Plotted results against baselines

Univariate Long Short-Term Memory

RMSE = 0.9593156724329953

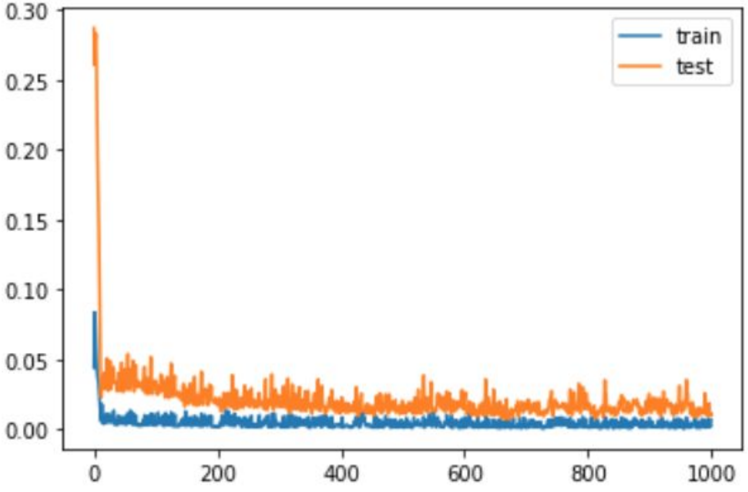


RMSE = 3.7566068880840042



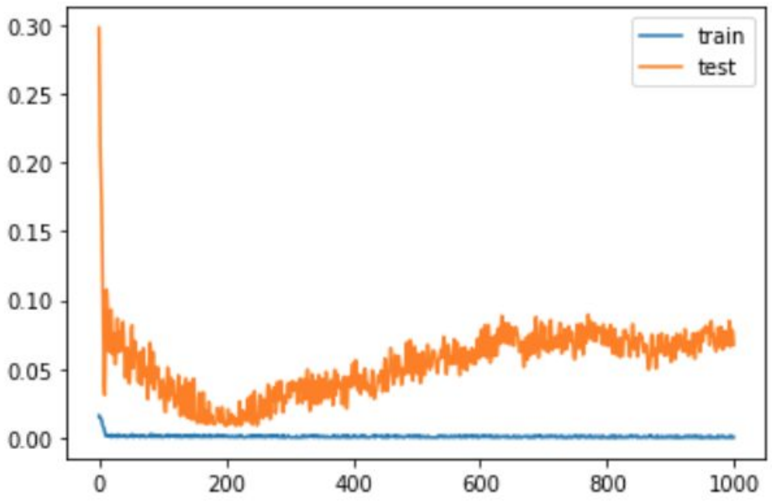
Multivariate Long Short-Term Memory

```
Epoch 1000/1000
8/8 - 0s - loss: 0.0066 - val_loss: 0.0108
```




```
RMSE =  1712.8453
```

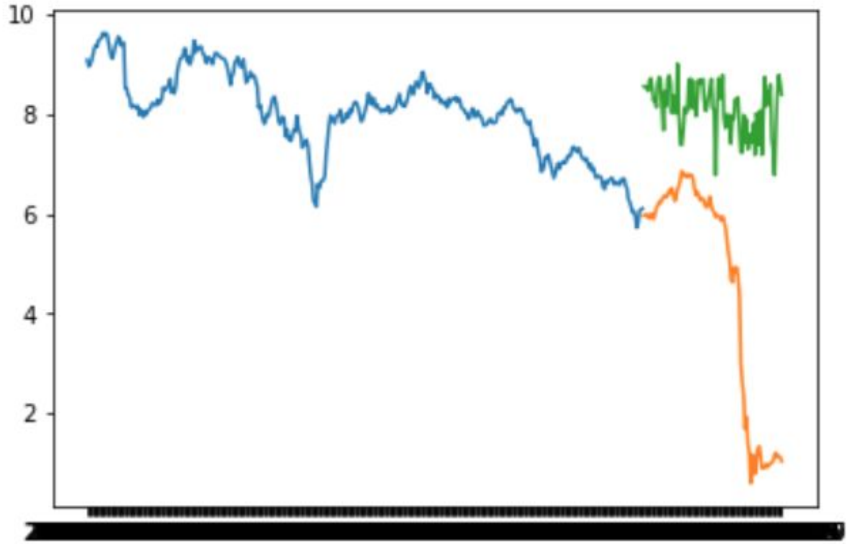
```
Epoch 1000/1000
8/8 - 0s - loss: 3.8514e-04 - val_loss: 0.0676
```



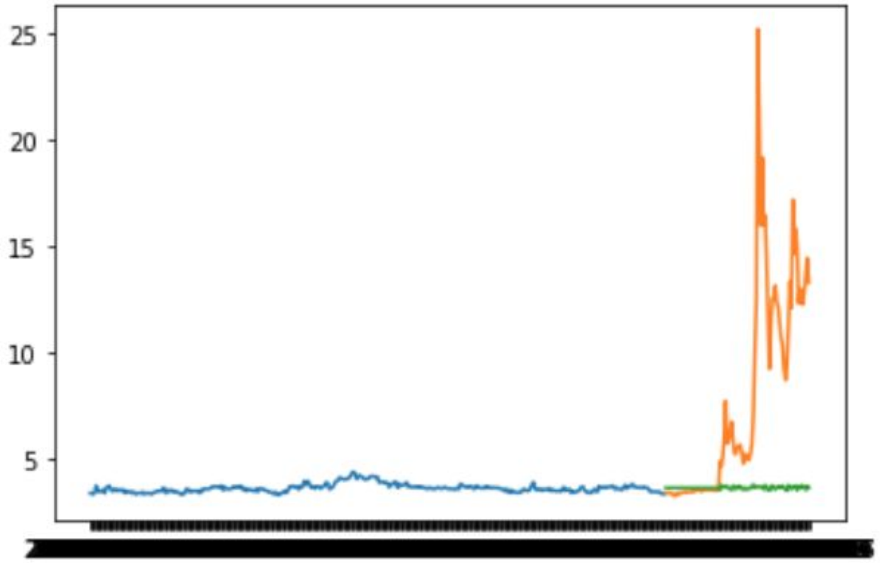

```
RMSE =  35630.48
```

K-Nearest Neighbor Using Price

RMSE = 4.101499186279391

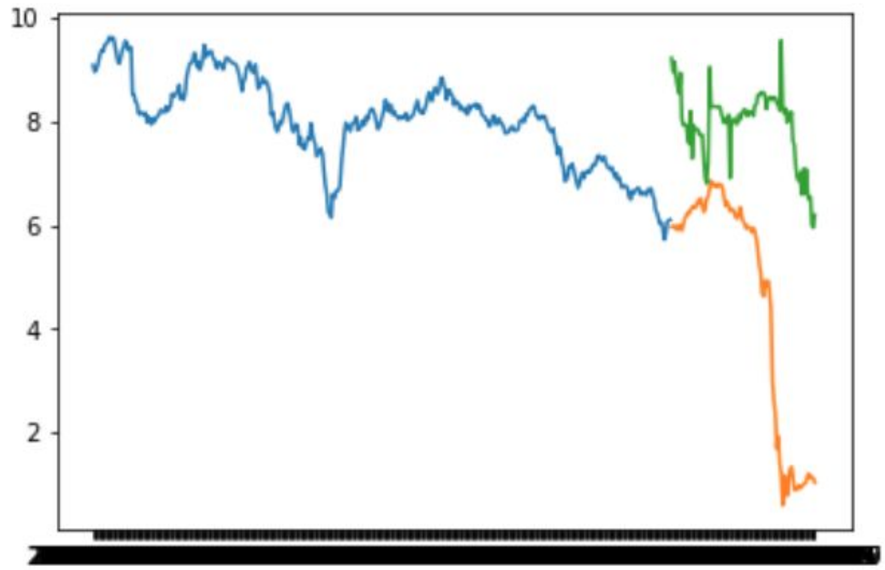


RMSE = 6.471844199668853

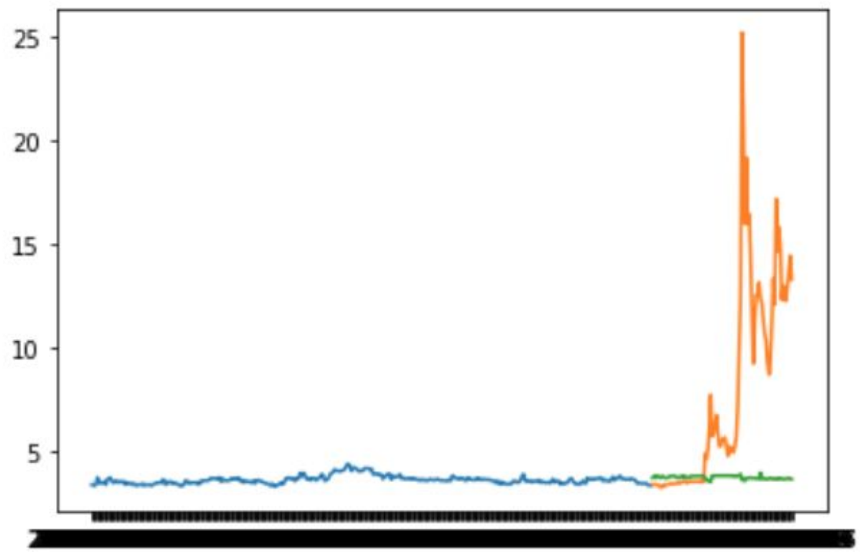


K-Nearest Neighbor Using Robinhood Data

RMSE = 3.9289020676965856



RMSE = 6.4469574741427



Conclusion

Tools

- SageMaker Jupyter Notebook
- EMR
- Pandas
- Sklearn
- Matplotlib

Impact

Discovered that Robinhood user popularity was not a good predictor of future price for stocks with high historical correlation with Robinhood popularity