Stock Prediction 2020

Prepared by: Kai Luo

Introduction

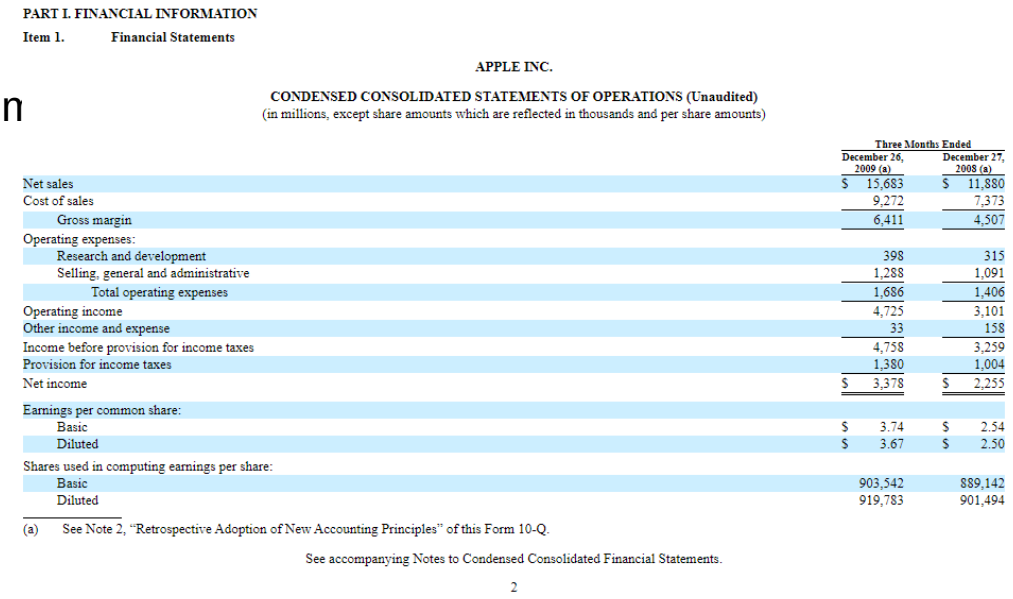
- Background
- Data Scraping
- Model 1: Text Data and Discrete Target
- Model 2: Numerical Data and Discrete Target
- Model 3: Numerical Data and Numerical Target
- Model 4: Time Series – Cyclic Stock
- Model 5: Time Series – Growth Stock
- Recommendations & Future Work
- Summary

Background

- This is an experiment to try to predict stock prices base on 10-Q quarterly statements.
- Stock History: https://github.com/ranaroussi/yfinance.
- Financial Statements: https://www.sec.gov/edgar/searchedgar/companysearch.html.
- 10 Years of S&P 500 Companies (current May 2021 list).

Data Scraping

- S&P 500: https://en.wikipedia.org/wiki/List_of_S%26P_500_companies
- EDGAR 10-Q Quarterly statements like these:
  - Methods are developed to extract “net incon
- Stock histories from 2010 to 2020 are obtained.



PART I. FINANCIAL INFORMATION

Item 1. Financial Statements

APPLE INC.

CONDENSED CONSOLIDATED STATEMENTS OF OPERATIONS (Unaudited)

(in millions, except share amounts which are reflected in thousands and per share amounts)

	Three Months Ended	
	December 26, 2009 (a)	December 27, 2008 (a)
Net sales	$ 15,683	$ 11,880
Cost of sales	9,272	7,373
Gross margin	6,411	4,507
Operating expenses:		
Research and development	398	315
Selling, general and administrative	1,288	1,091
Total operating expenses	1,686	1,406
Operating income	4,725	3,101
Other income and expense	33	158
Income before provision for income taxes	4,758	3,259
Provision for income taxes	1,380	1,004
Net income	$ 3,378	$ 2,255
Earnings per common share:		
Basic	$ 3.74	$ 2.54
Diluted	$ 3.67	$ 2.50
Shares used in computing earnings per share:		
Basic	903,542	889,142
Diluted	919,783	901,494

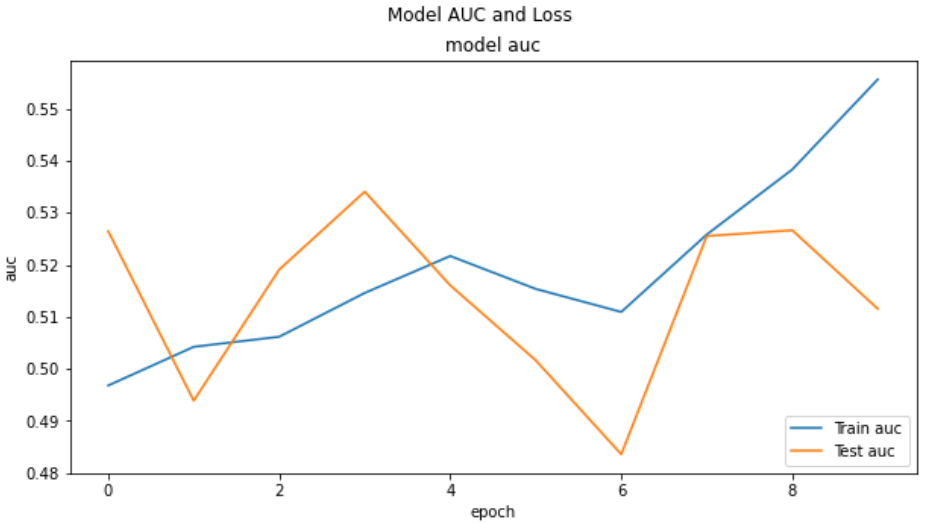
(a) See Note 2, “Retrospective Adoption of New Accounting Principles” of this Form 10-Q.

See accompanying Notes to Condensed Consolidated Financial Statements.

2

Model 1: Text Data and Discrete Target

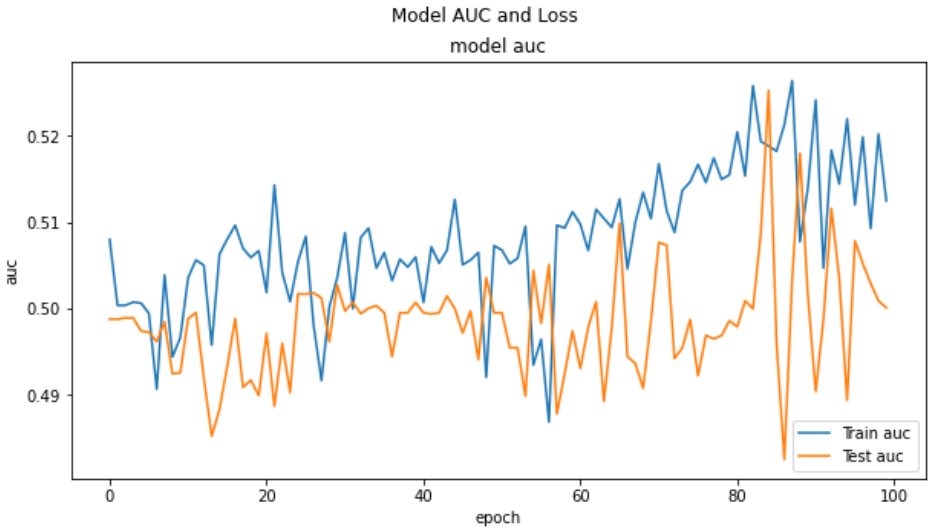
- Input: scraped text word stems with numbers.
  - Humans can read it to compare magnitudes, this was a test to see check if NLP can do numerical reasons as well.
- Target: Has the stock increased by 2% or more in a quarter?
- No correlation.



	Symbol	target	Raw_Increase	Net_Income	content
0	A	False	-0.033323	[79.0, 64.0]	net incom 79 64 denomin basic weight averag sh...
1	AAPL	True	0.160159	[3378.0, 2255.0]	cost sale 9272 7373 gross margin 6411 4507 ope...
2	ABC	True	0.172716	[151307.0, 111056.0]	revenu 19335859 17338377 cost good sold 187724...
3	ABMD	False	-0.058937	[461879.0, 502408.0]	revenu reimburs net revenu 5176438 5266324 105...
4	ACN	False	-0.085217	[256.8, 73.6]	sale 2173 5 2195 3 cost sale 1568 6 1629 7 sel...

Model 2: Numerical Data and Discrete Target

- Input: [Current Quarter Net Income, Past Quarter Net Income, Abs Difference, Frac Difference.].
- Target: Has the stock increased by 2% or more in a quarter?
- No correlation.

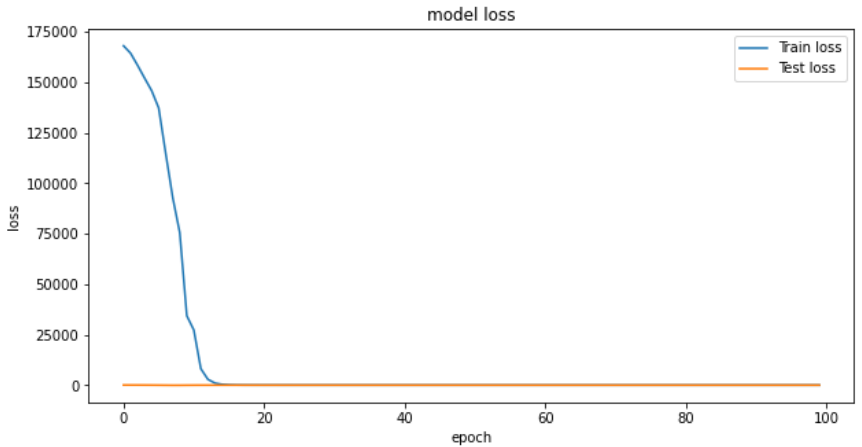


	content	target
0	[79.0, 64.0, 15.0, 0.234009360374415]	False
1	[3378.0, 2255.0, 1123.0, 0.4979823511152499]	True
2	[151307.0, 111056.0, 40251.0, 0.3624384432732646]	True
3	[461879.0, 502408.0, -40529.0, -0.080669479652...	False
4	[256.8, 73.6, 183.20000000000002, 2.4857530529...	False
5	[93330.0, 75963.0, 17367.0, 0.22862416094129911]	False
6	[123333.0, 115864.0, 7469.0, 0.06446345330434534]	False
7	[316376.0, 312068.0, 4308.0, 0.01380467917098864]	False
8	[357549.0, 109929.0, 247620.0, 2.2525427752978...	True
9	[239.2, 168.5, 70.69999999999999, 0.4193357058...	False
10	[20423.0, 23873.0, -3450.0, -0.1445141184010455 3]	False
11	[565.0, 10.0, 555.0, 54.95049504950495]	True
12	[71005.0, 64407.0, 6598.0, 0.10244212206418236]	False

Model 3: Numerical Data and Numerical Target

- Input: [Current Quarter Net Income, Past Quarter Net Income, Abs Difference, Frac Difference.].
- Target: Stock Price
- No correlation.
- Loss = Abs Percentage Error.

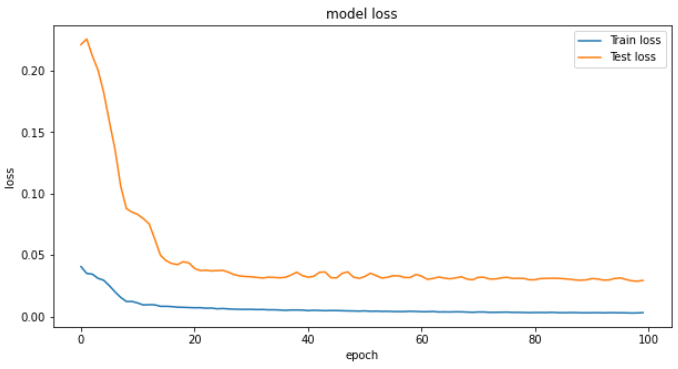
```
Epoch 96/100
3/3 - 0s - loss: 104.2151 - val_loss: 99.9939
Epoch 97/100
3/3 - 0s - loss: 104.1951 - val_loss: 99.9939
Epoch 98/100
3/3 - 0s - loss: 104.1488 - val_loss: 99.9940
Epoch 99/100
3/3 - 0s - loss: 104.1305 - val_loss: 99.9940
Epoch 100/100
3/3 - 0s - loss: 104.1105 - val_loss: 99.9941
Wall time: 3.88 s
```



	content	target
0	[79.0, 64.0, 15.0, 0.234009360374415]	-0.033323
1	[3378.0, 2255.0, 1123.0, 0.4979823511152499]	0.160159
2	[151307.0, 111056.0, 40251.0, 0.3624384432732646]	0.172716
3	[461879.0, 502408.0, -40529.0, -0.080669479652...	-0.058937
4	[256.8, 73.6, 183.20000000000002, 2.4857530529...	-0.085217
5	[93330.0, 75963.0, 17367.0, 0.22862416094129911]	-0.061227
6	[123333.0, 115864.0, 7469.0, 0.06446345330434534]	-0.028350
7	[316376.0, 312068.0, 4308.0, 0.01380467917098864]	-0.190939
8	[357549.0, 109929.0, 247620.0, 2.2525427752978...	0.090218
9	[239.2, 168.5, 70.69999999999999, 0.4193357058...	-0.045543
10	[20423.0, 23873.0, -3450.0, -0.14451411840104553]	-0.010719
11	[565.0, 10.0, 555.0, 54.95049504950495]	0.149290
12	[71005.0, 64407.0, 6598.0, 0.10244212206418236]	-0.023239

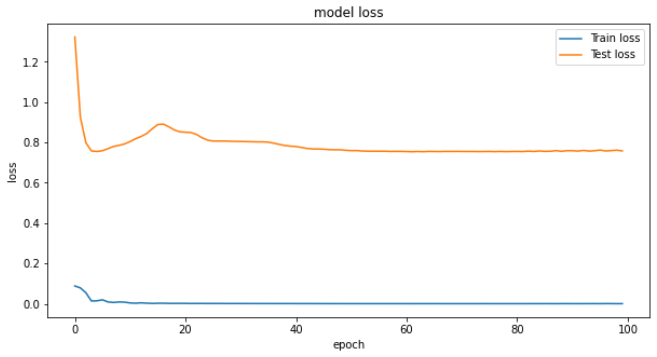
Model 4: Time Series – Cyclic Stock

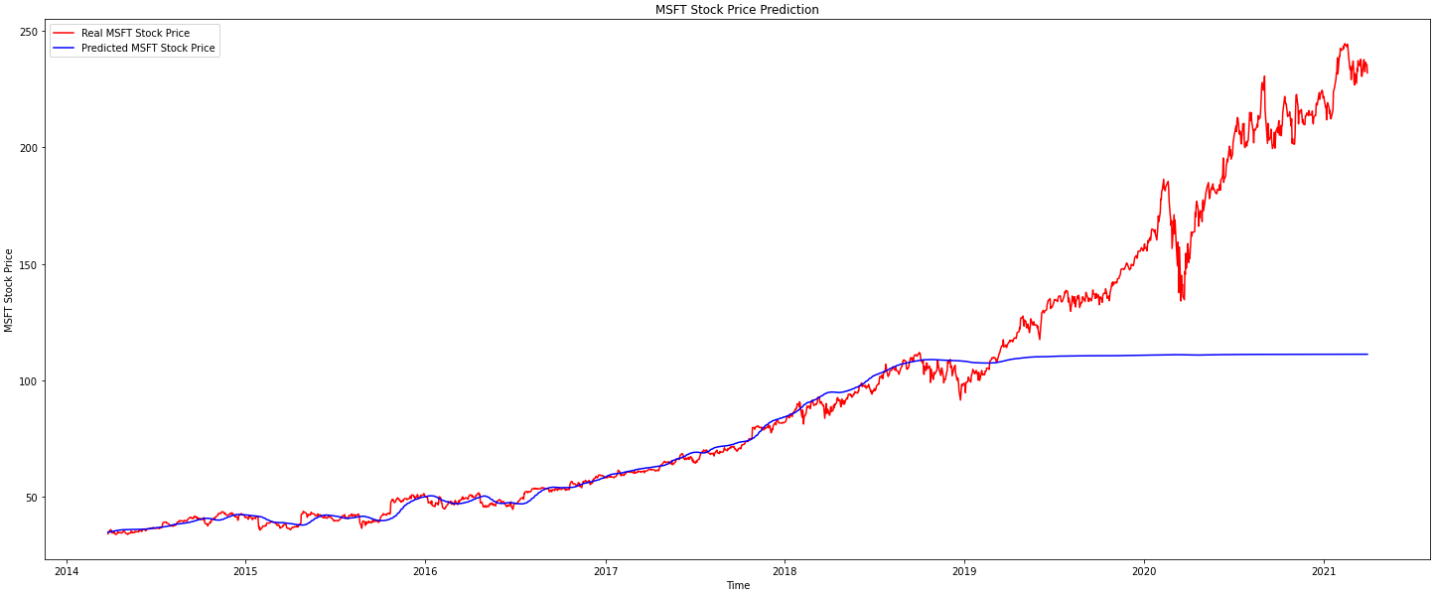
- Input: 60 days of past stock prices for Ford.
- Target: tomorrow's stock price.
- Right shifted.
- Loss = Mean Squared Error.



Model 5: Time Series – Growth Stock

- Input: 60 days of past stock prices for Microsoft.
- Target: tomorrow's stock price.
- Right shifted, and extremes out cut off.
- Loss = Mean Squared Error.





Recommendations and Future Works

1) Time Series Models may be useful for cyclic stocks, but do not trust them as soon has there are new highs or lows.
2) Develop better scraping techniques for the Statements.
3) Try developing models based on sentiments to see if that works better for stock price movement prediction.

Summary

- Scraped company quarterly statements from EDGAR and stock histories from yfinance.
- Made 5 different models to try to predict stock prices:
  - Model 1: Text Data and Discrete Target
  - Model 2: Numerical Data and Discrete Target
  - Model 3: Numerical Data and Numerical Target
  - Model 4: Time Series – Cyclic Stock
  - Model 5: Time Series – Growth Stock
- Conclusions:
  - Stock movement prediction is hard and mostly inaccurate.
- Therefore:
- Time Series Models may be useful for cyclic stocks, but do not trust them as soon has there are new highs or lows.
- Develop better scraping techniques for the Statements.
- Try developing models based on sentiments to see if that works better for stock price movement prediction.

Thank you for you attention!

Any questions?