Stock Market Project

Lehman Brothers Risk Management Department

Alessandro Ferrari, Rishi Bhargava, Sergio Estefano, Rakshit Kaushik

Introduction of Data

- Past stock data obtained using the quantmod package
  - Columns: Opening prices, high prices, low prices, closing prices, and the adjusted prices
  - Rows: Date (Data is taken from January 3, 2007 to the current date)
- Calculated Data
  - Moving averages (short and long term) found using TTR library
  - Date of golden crosses found using detect-cross function
  - Price difference between moving averages found using MovingAvgDiff function
- Data sorted obtained and sorted into data frame using getCorrectData function
- Data accessed using findIndex function
- Dimensions
  - Number of rows: 1641
  - Number of columns: 19 per each stock

Full Data Frame

	AAPL.High	AAPL.Open	AAPL.Close	AAPL.Low	AAPL.Volume	AAPL.Adjusted	AAPL.High.SMA50	AAPL.Open.SMA50	AAPL.Close.SMA50	AAPL.Low.SMA50	AAPL.Volume.SMA50	AAPL.Adjusted.SMA50	AAPL.High.SMA200	AAPL.Open.SMA200	AAPL.Close.SMA200	AAPL.Low.SMA200	AAPL.Volume.SMA200	AAPL.Adjusted.SMA200	Crosses
2016-08-29	26.8600	26.6550	26.7050	26.5725	99881200	25.08255	25.57615	25.39780	25.43510	25.27065	130113016	23.81038	25.89656	25.67132	25.67004	25.44159	159636600	23.86712	NA
2016-08-30	26.6250	26.4500	26.5000	26.3750	99455600	24.89001	25.62580	25.44680	25.48960	25.32300	129349176	23.86397	25.88366	25.65825	25.65789	25.42890	159483366	23.85798	NA
2016-08-31	26.6425	26.4150	26.5250	26.4100	118649600	24.91349	25.67690	25.50040	25.54055	25.37780	128878456	23.91425	25.87241	25.64633	25.65009	25.42061	159160366	23.85286	NA
2016-09-01	26.7000	26.5350	26.6825	26.4050	106806000	25.06142	25.72645	25.54985	25.59645	25.42915	128677048	23.96917	25.86311	25.63978	25.64078	25.41389	158932262	23.84636	NA
2016-09-02	27.0000	26.9250	26.9325	26.7050	107210000	25.29623	25.78500	25.60865	25.65460	25.48700	128242032	24.02622	25.85430	25.63075	25.63333	25.40576	158915974	23.84160	g
2016-09-06	27.0750	26.9750	26.9250	26.8775	107521600	25.28919	25.85320	25.68360	25.72610	25.56130	124367552	24.09574	25.84281	25.62093	25.62134	25.39578	158520088	23.83264	NA
2016-09-07	27.1900	26.9575	27.0900	26.7675	169457200	25.44416	25.93175	25.75775	25.80770	25.63915	124117528	24.17471	25.82908	25.60866	25.60831	25.38366	158501458	23.82274	NA
2016-09-08	26.8175	26.8125	26.3800	26.3100	212008000	24.77730	25.99980	25.82950	25.86735	25.70465	125122096	24.23311	25.81326	25.59373	25.59109	25.36665	158875756	23.80891	NA
2016-09-09	26.4300	26.1600	25.7825	25.7825	186228000	24.21610	26.05565	25.88285	25.91100	25.75215	125924176	24.27649	25.79575	25.57544	25.57281	25.34889	159157246	23.79406	NA
2016-09-12	26.4300	25.6625	26.3600	25.6325	181171200	24.75852	26.10540	25.92390	25.96020	25.79330	126680688	24.32512	25.77871	25.55709	25.55601	25.33065	159207038	23.78062	NA
2016-09-13	27.1975	26.8775	26.9875	26.8100	248704800	25.34789	26.16700	25.98400	26.02050	25.85285	129572664	24.38419	25.76566	25.54246	25.54341	25.31730	160022796	23.77110	NA
2016-09-14	28.2575	27.1825	27.9425	27.1500	443554800	26.24487	26.25515	26.05070	26.10440	25.92355	136227344	24.46539	25.75894	25.53051	25.53586	25.30605	161979642	23.76633	NA
2016-09-15	28.9325	28.4650	28.8925	28.3725	359934400	27.13716	26.35550	26.14700	26.20460	26.01915	140950104	24.56192	25.75434	25.52535	25.53245	25.30073	162995708	23.76545	NA
2016-09-16	29.0325	28.7800	28.7300	28.5100	319547600	26.98453	26.45365	26.24410	26.29950	26.11125	145329888	24.65348	25.75099	25.52081	25.52943	25.29720	163896398	23.76491	NA
2016-09-19	29.0450	28.7975	28.3950	28.3125	188092000	26.66988	26.55010	26.33760	26.38400	26.19725	146778760	24.73530	25.74858	25.51813	25.52605	25.29366	164169126	23.76403	NA
2016-09-20	28.5300	28.2625	28.3925	28.1275	138057200	26.66753	26.63245	26.41910	26.46695	26.27615	147636312	24.81566	25.74524	25.51375	25.52401	25.29153	164028022	23.76438	NA
2016-09-21	28.4975	28.4625	28.3875	28.1100	144012800	26.66283	26.71390	26.50250	26.54760	26.35275	148583168	24.89388	25.73866	25.51195	25.51716	25.28819	163592546	23.76028	NA
2016-09-22	28.7350	28.5875	28.6550	28.5000	124296000	26.91409	26.80025	26.58720	26.63635	26.43855	148997712	24.97969	25.73251	25.50616	25.51259	25.28343	163572342	23.75831	NA
2016-09-23	28.6975	28.6050	28.1775	27.8875	209924800	26.46559	26.87925	26.67235	26.70595	26.50970	150082688	25.04756	25.72775	25.50229	25.50569	25.27679	163935776	23.75416	NA
2016-09-26	28.3475	27.9100	28.2200	27.8875	119477600	26.50551	26.94970	26.73595	26.77645	26.57495	150061280	25.11627	25.72238	25.49479	25.50226	25.27238	163605936	23.75321	NA
2016-09-27	28.2950	28.2500	28.2725	28.0850	98429600	26.55483	27.01495	26.80745	26.84275	26.64365	149110360	25.18107	25.71768	25.49099	25.49841	25.26841	163513830	23.75188	NA
2016-09-28	28.6600	28.4225	28.4875	28.3575	118564400	26.75676	27.08815	26.87810	26.91315	26.71410	149579256	25.24972	25.71674	25.48911	25.49938	25.26914	163168928	23.75501	NA
2016-09-29	28.4500	28.2900	28.0450	27.9500	143548000	26.34115	27.15485	26.94390	26.97425	26.77440	150348136	25.30964	25.71814	25.49034	25.49900	25.27165	162600294	23.75687	NA
2016-09-30	28.3425	28.1150	28.2625	27.9500	145516400	26.54543	27.21670	27.00705	27.04235	26.83775	150642304	25.37612	25.71885	25.49099	25.50220	25.27346	162261414	23.76204	NA
2016-10-03	28.2625	28.1775	28.1300	28.0700	86807200	26.42098	27.28545	27.07430	27.11165	26.90760	150113352	25.44370	25.72018	25.49304	25.50368	25.27781	161570680	23.76562	NA
2016-10-04	28.5775	28.2650	28.2500	28.1575	118947200	26.53369	27.36280	27.14835	27.18995	26.98615	149261664	25.51971	25.72275	25.49434	25.50870	25.28238	161269960	23.77248	NA
2016-10-05	28.4150	28.3500	28.2625	28.1725	85812400	26.54543	27.44125	27.23125	27.27185	27.06750	146478728	25.59908	25.72793	25.49995	25.51748	25.29098	159769956	23.78281	NA
2016-10-06	28.5850	28.4250	28.4725	28.2825	115117200	26.74267	27.49120	27.27840	27.32655	27.11940	141393488	25.65306	25.73664	25.50798	25.52568	25.30043	159393730	23.79262	NA
2016-10-07	28.6400	28.5775	28.5150	28.3775	97433600	26.78258	27.54175	27.33580	27.37515	27.17285	140152576	25.70135	25.74519	25.51661	25.53421	25.30925	159225110	23.80274	NA
2016-10-10	29.1875	28.7550	29.0125	28.6800	144944000	27.24987	27.60275	27.38995	27.43435	27.22805	140832760	25.75959	25.75506	25.52630	25.54351	25.31865	159296682	23.81361	NA
2016-10-11	29.6725	29.4250	29.0750	29.0500	256164000	27.30857	27.66545	27.45640	27.48560	27.28700	142902608	25.81041	25.76718	25.53718	25.55385	25.32896	160306094	23.82545	NA
2016-10-12	29.4950	29.3375	29.3350	29.1875	150347200	27.55277	27.72500	27.51290	27.54990	27.35075	143204224	25.87345	25.78004	25.54938	25.56700	25.34218	160523746	23.83990	NA
2016-10-13	29.3600	29.1975	29.2450	28.9300	140769600	27.46824	27.78300	27.57280	27.60585	27.40550	143603408	25.92867	25.79005	25.56166	25.57730	25.35325	160608970	23.85171	NA
2016-10-14	29.5425	29.4700	29.4075	29.2825	142608800	27.62086	27.84385	27.63430	27.66465	27.46475	144262888	25.98390	25.80189	25.57329	25.59019	25.36569	160817738	23.86592	NA
2016-10-17	29.4600	29.3325	29.3875	29.1950	94499600	27.60208	27.89480	27.68960	27.71500	27.51775	142908608	26.03119	25.81540	25.58619	25.60555	25.38064	160471990	23.88242	NA
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

view data on github

Introduction of Algorithm

Model 1: Linear Regression

$$y = \alpha + \beta x + \epsilon$$

x = time, y = stock price, = y intercept, $\epsilon = error$

- Linear regression is used to find the relationship between two variables, or in our case, x as y stated above
- Linear not optimal for predicting stocks, any sudden change in price can cause a user to lose money

Model 2: Golden Cross RNN

- RNN's are designed to predict stock data; a golden cross occurs when the plotted line of a stock's long term average crosses the line of its short term average.
- If the short term average starts below the long term average and crosses above it, the pattern is called a golden cross. Otherwise, it's called a death cross.

Introduction of Algorithm CONT.

Model 3: Stock Return RNN

- Stock price trends vary from year to year, so training an AI to predict next year's stock closing prices using last year's closing price data is un-ideal. Stock returns don't have as much variation and are better suited for making predictions with an RNN.

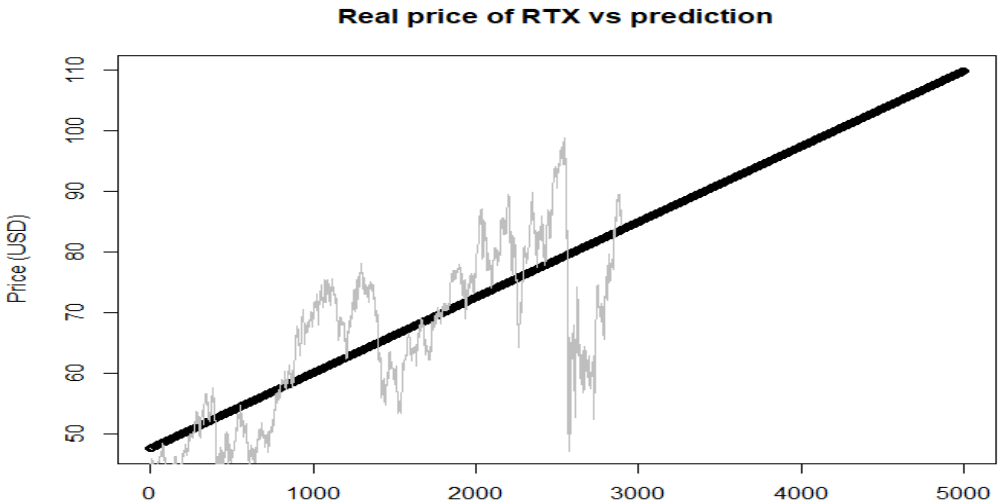
Photo of golden cross/death cross

Algorithm Results

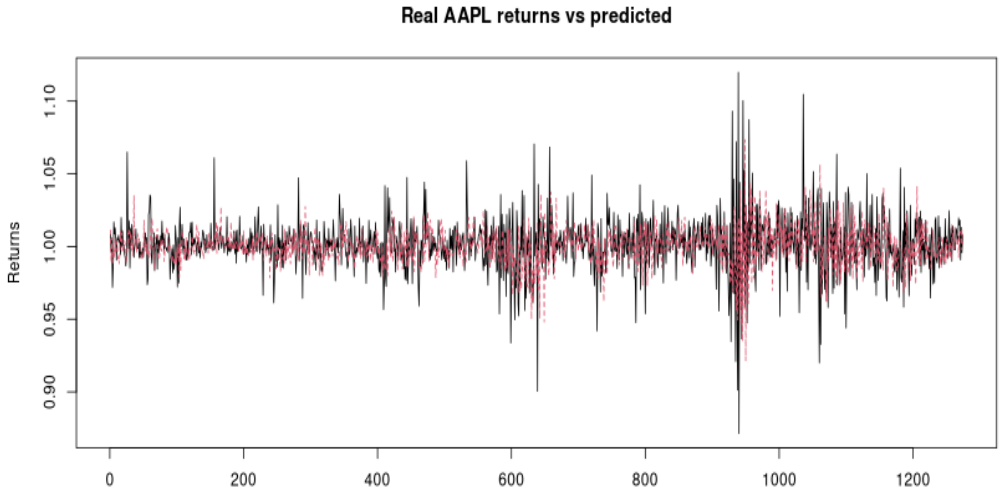
Error (Tested on AAPL with 100-150 Epochs where applicable):

- Linear Regression: $\pm 30\%$
- Golden Cross RNN: $\pm 35\%$
- Return RNN: $\pm 1\%$
- Price RNN: $\pm 14\%$

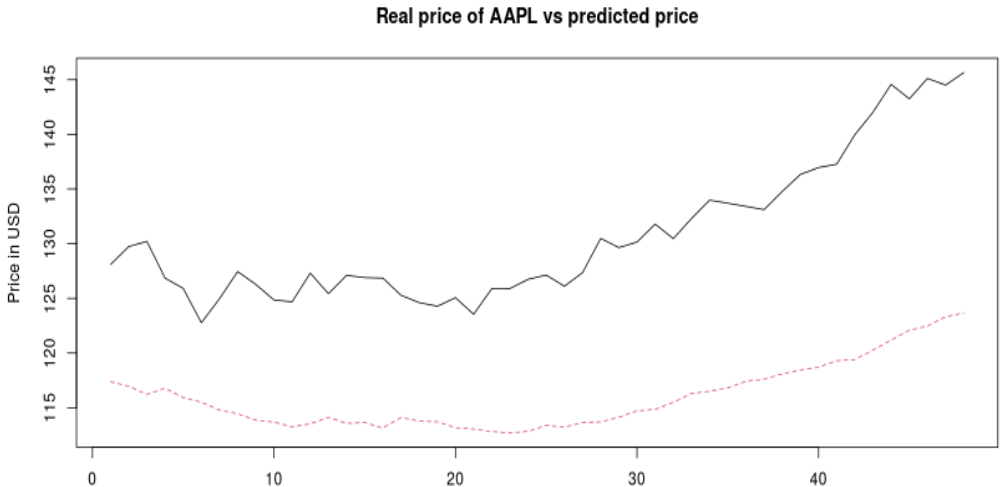
RTX prediction v Real Price



AAPL Returns Prediction v Real Price



AAPL Price Prediction v Real Price



Conclusions

- Research question: ”What model would be best for analyzing different machine learning algorithms and observe the viability of machine learning models to predict the stock market?”
- Linear Regression (m1) : This model was not the most optimal because we were trying to predict stocks and not just compare two variables.
- Stock Return (m2): The stock return model was solid as designed with RNN’s which are made for prediction problems. The neural network used stock returns as both the explanatory and response variables. The stock return model doesn’t have as much variation and are better suited for making predictions with an RNN.
- Golden Cross RNN (m3): The last model, RNN using golden crosses, th explanatory variable is price data leading up to a golden crosses, and the response variable is price data after the golden cross. This model attempted to predict the stock price outcome after a golden cross.

The End

Thank You!