

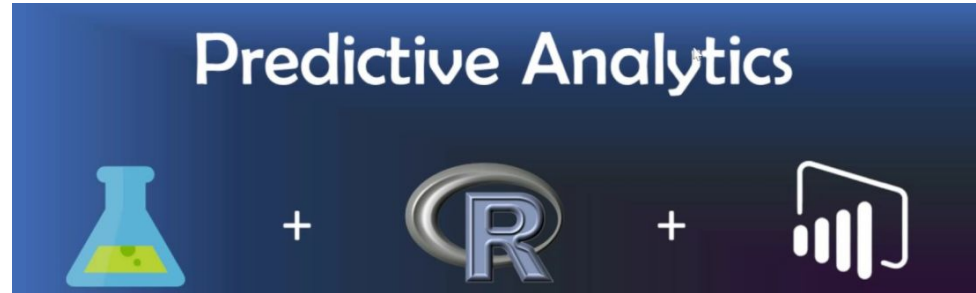
# MAXIMIZING PORTFOLIO RETURNS USING MACHINE LEARNING

## TEAM 8



Harsh Pathak  
Lei Li  
Krushika Tapedia  
Mukund Khandelwal

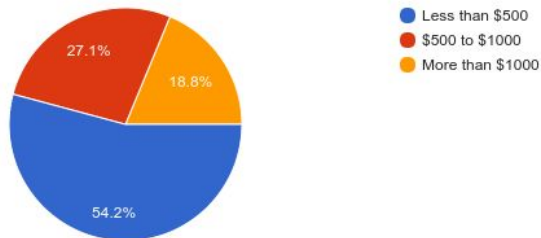
# Motivation



# Survey

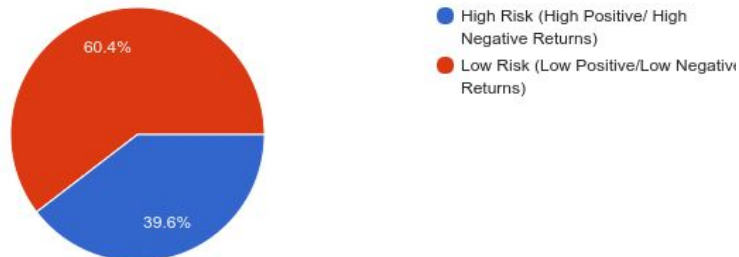
If you decide to invest in the financial markets, how much are you initially willing to invest?

48 responses



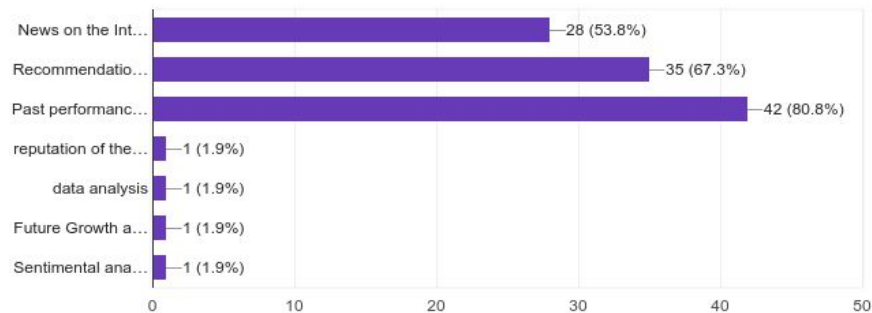
Risk Tolerance?

48 responses



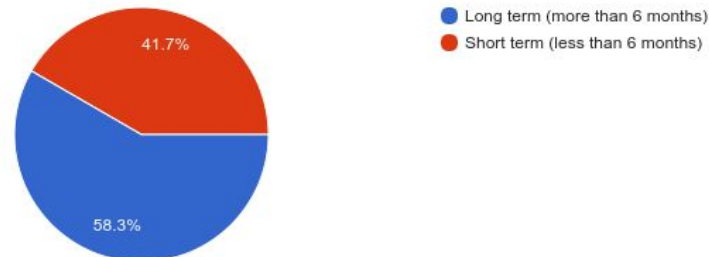
When making investment decisions, which of the following factors will you consider most important? Choose one or more than one option:

52 responses



For how long do you plan to invest? (options relative to our project scope)

48 responses



# Introduction

---



 ZERODHA

 DataCamp

 INVESTOPEDIA

 -bloggers

**Trading Geeks**

# Stocks Chosen



Correlation Matrix :

	Bank of America	Microsoft	ANI Pharmaceuticals
Bank of America	<i>1.00</i>	<i>0.3958</i>	<i>0.2740</i>
Microsoft	<i>0.3958</i>	<i>1.00</i>	<i>0.2301</i>
ANI Pharmaceuticals	<i>0.2740</i>	<i>0.2301</i>	<i>1.00</i>

**Bank of America**



# Statistical Methods Applied :



- 1) Feature Engineering.
- 2) ARIMA.
- 3) Simple Linear Regression.
- 4) Multiple Linear Regression.
- 5) Principal Component Analysis.
- 6) Random Forests.
- 7) Splines.
- 8) General Additive Model.



# Feature Engineering : TA Analysis



# Curious Case of 'X' and 'Y'



X (Close)	Y(Tomorrow's Close)
100	102
102	99
99	105
105	?

X(Yesterday's Close)	Y(Close)
100	100
100	102
102	99
99	?

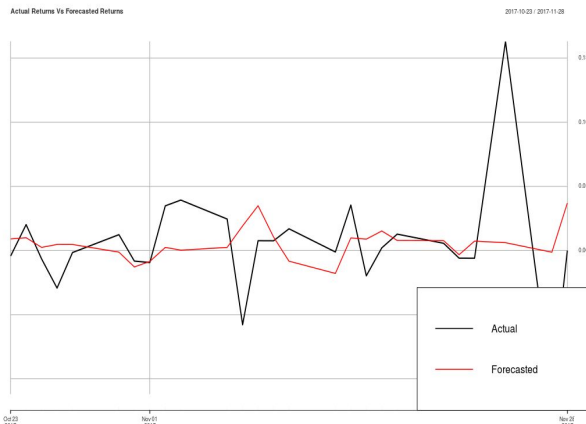
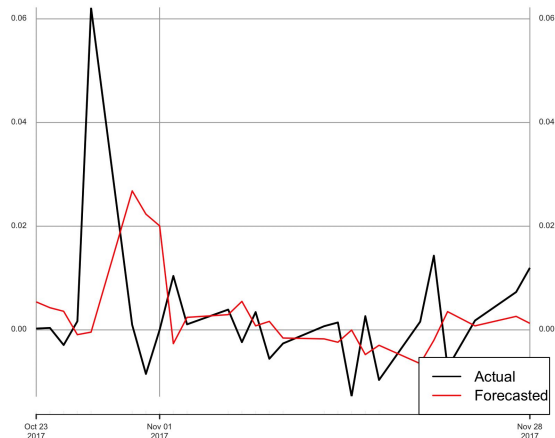


# ARIMA



**MSFT:**  
**Accuracy: 46.15%**

Results using Arima for MSFT: Actual Returns Vs Forecasted Returns 2017-10-23 / 2017-11-28

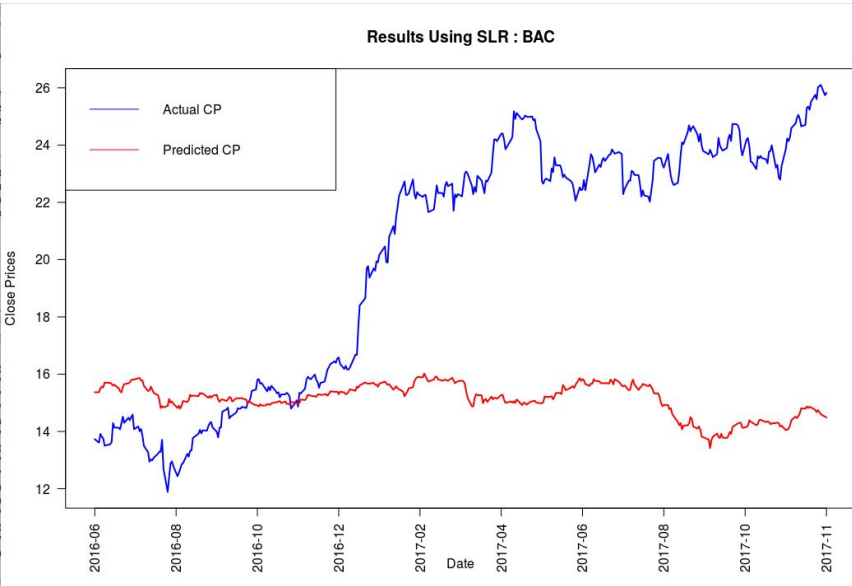


**ANIP:**  
**Accuracy: 61.53%**

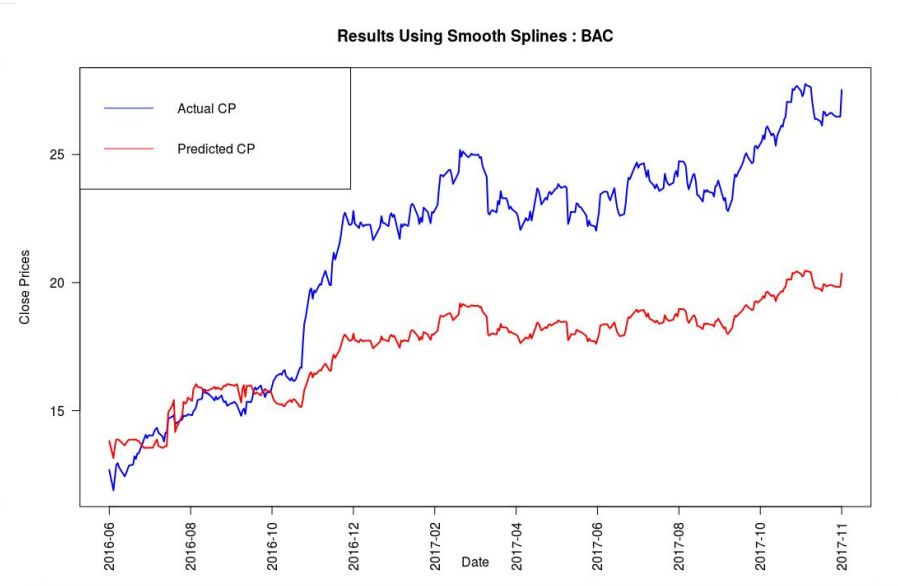
**BOA:**  
**Accuracy: 38.46%**



# SLR vs Splines : Bank of America



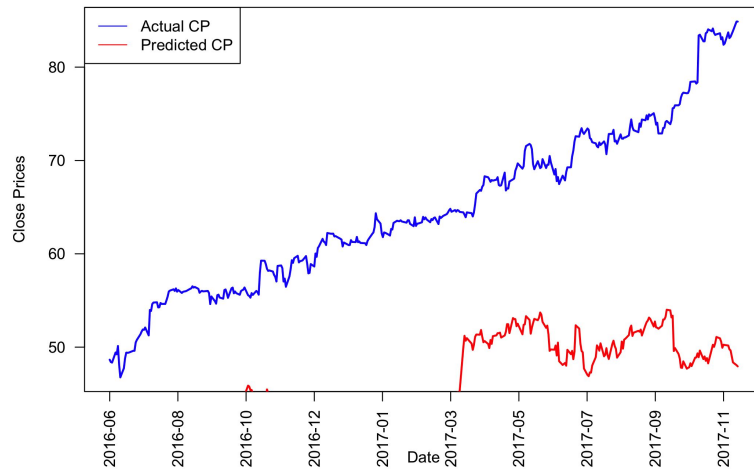
RMSE : 7.59



RMSE : 4.55

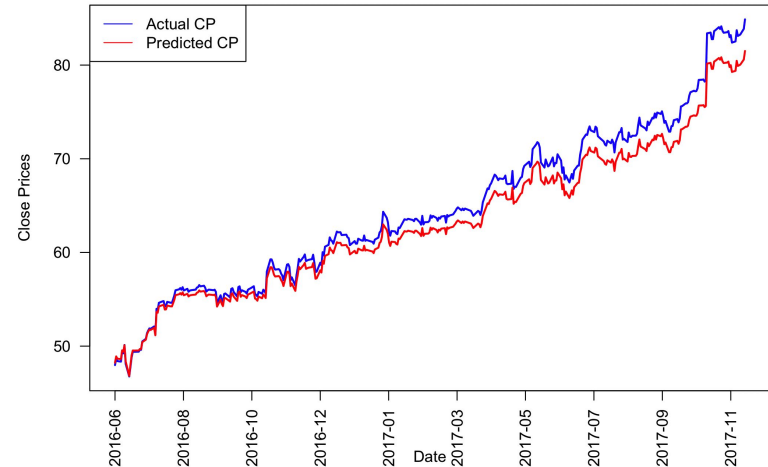
# SLR vs Splines : Microsoft

Results Using SLR : MSFT



**RMSE: 19.8608**

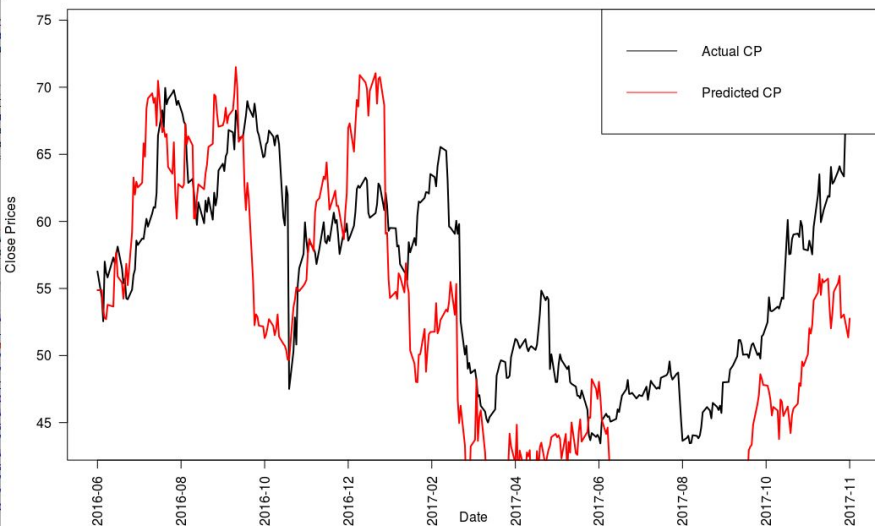
Results Using Splines : MSFT



**RMSE: 1.64362**

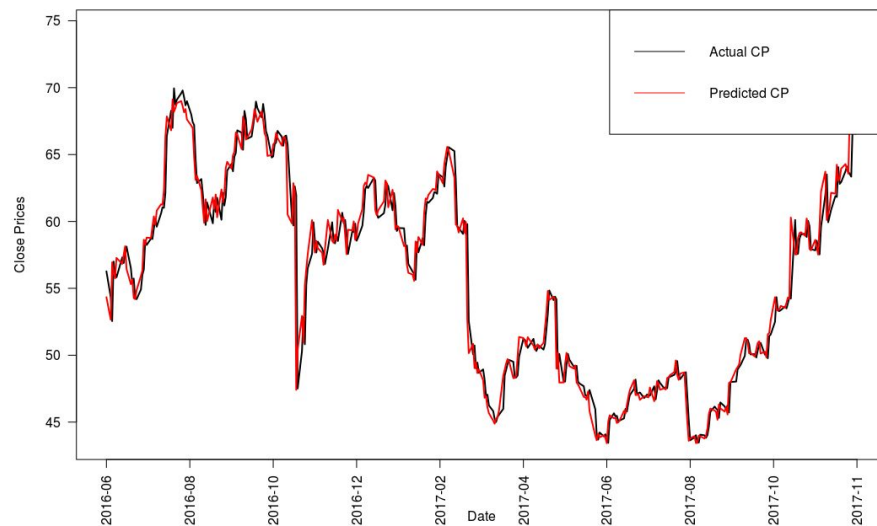
# SLR vs Splines : ANI Pharmaceuticals

Results Using SLR



**RMSE:8.8**

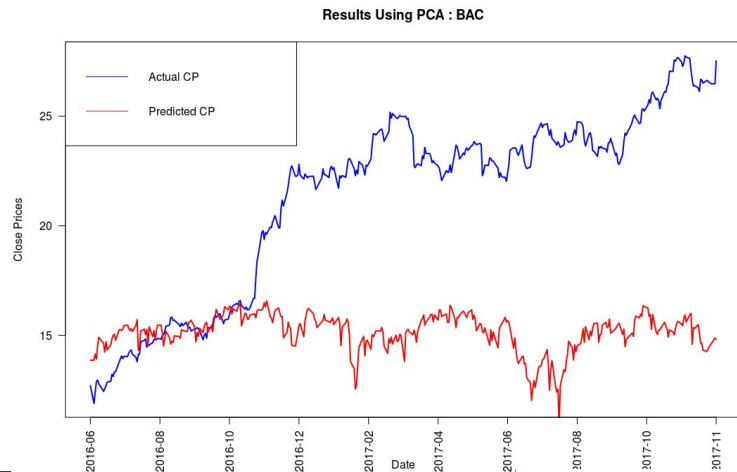
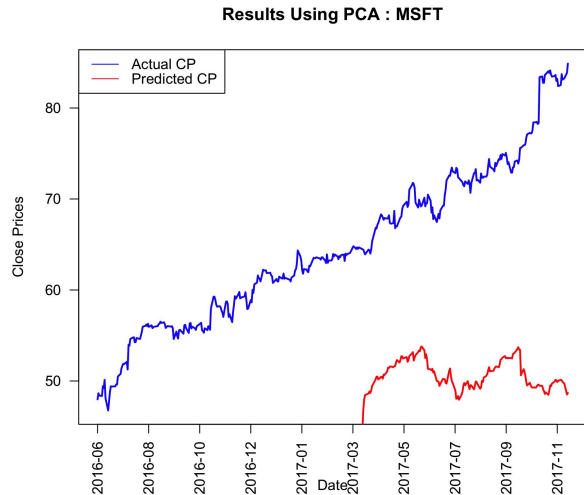
Results Using Splines : ANIP



**RMSE: 1.65**

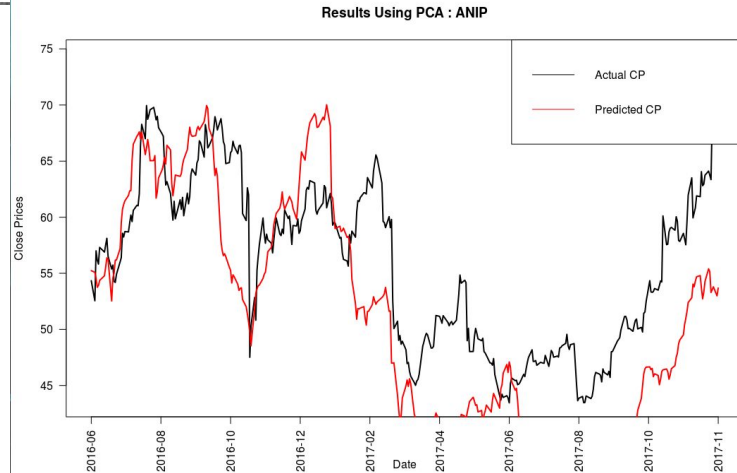
# Dimensionality Reduction

**RMSE : 19.82**



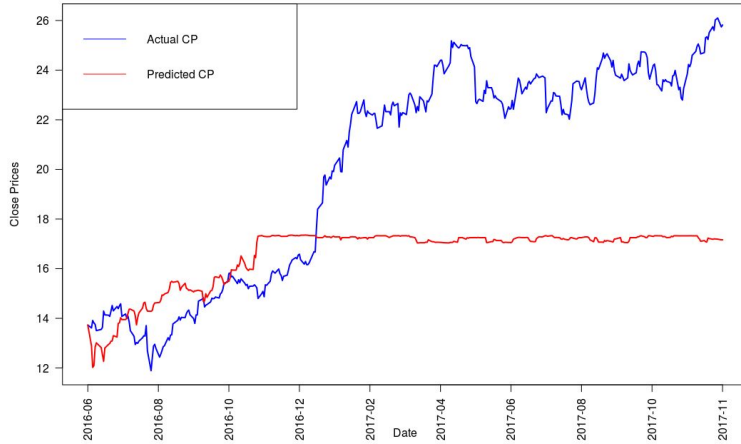
**RMSE : 7.59**

**RMSE : 8.48**

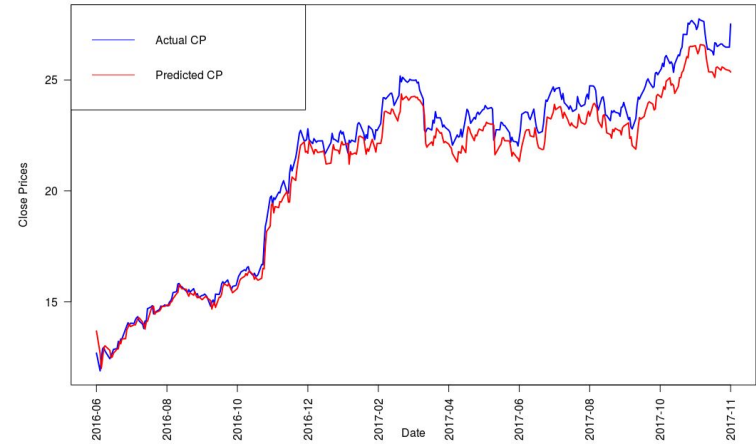


# MLR vs Random Forests vs GAM : Bank of America

Results Using Random Forests : BAC

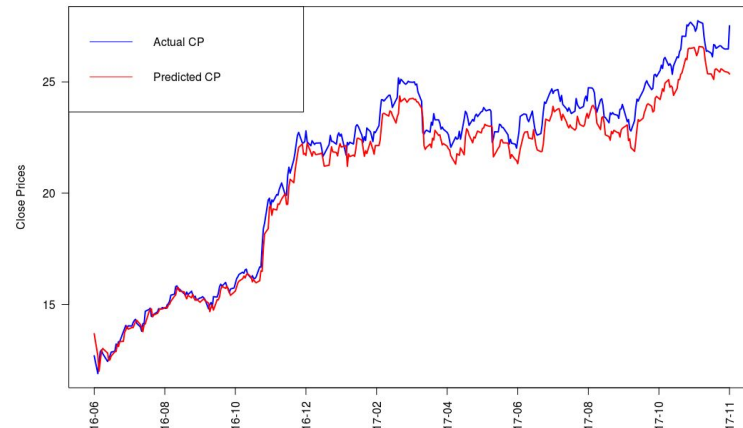


Results Using MLR : BAC



**RMSE : 5.69**

Results Using GAM : BAC



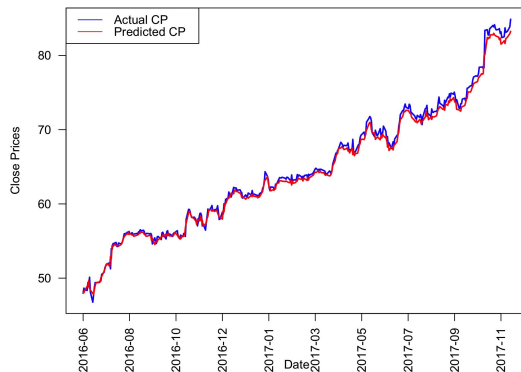
**RMSE : 0.76**

**RMSE : 0.76**



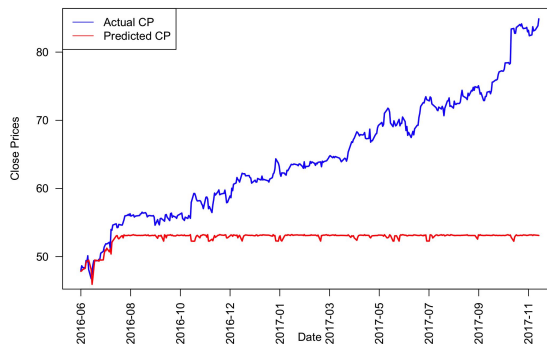
# MLR vs Random Forests vs GAM : Microsoft

Results Using MLR : MSFT



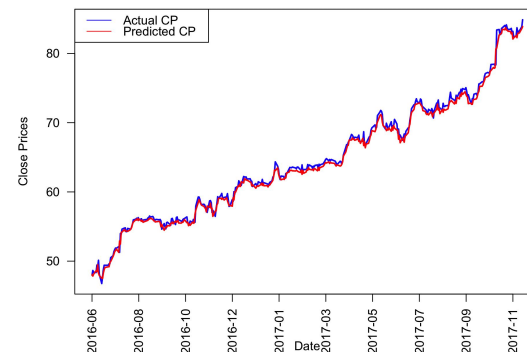
**RMSE = 0.70628**

Results Using Random Forests : MSFT



**RMSE = 14.66471**

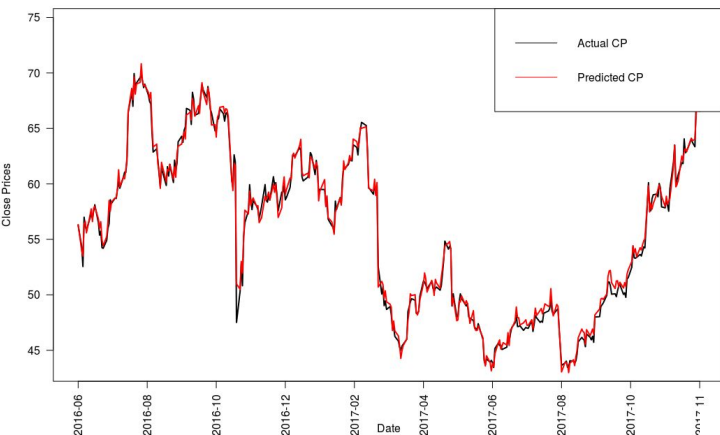
Results Using GAM : MSFT



**RMSE = 0.59195**

# MLR vs Random Forests vs GAM : ANI Pharmaceuticals

Results Using MLR : ANIP

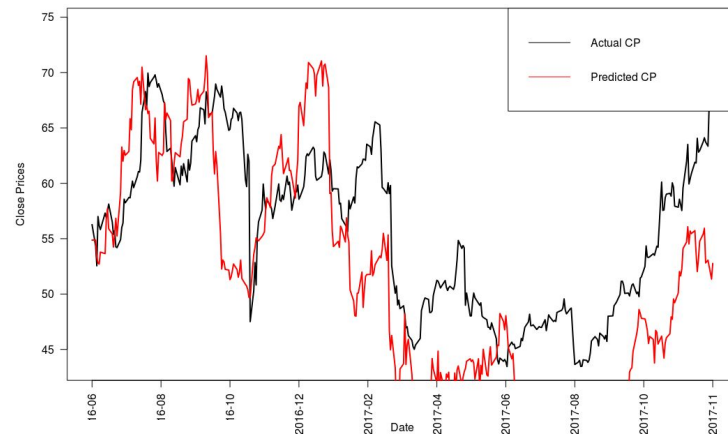


**RMSE : 1.64**

**RMSE : 1.75**

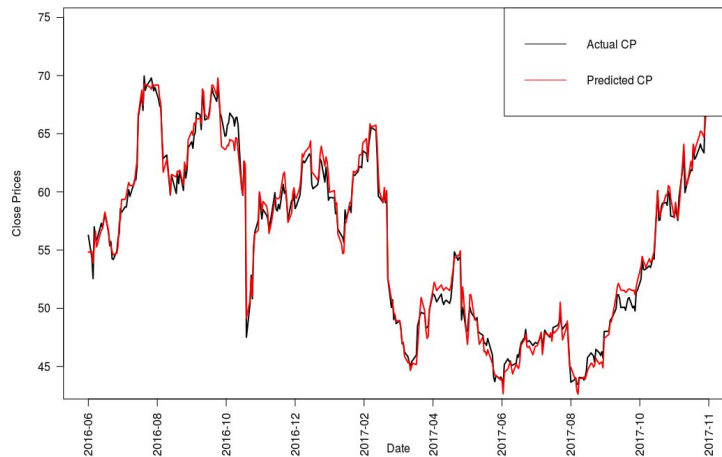


Results Using GAM : ANIP



**RMSE : 2.11**

Results Using Random Forests : ANIP



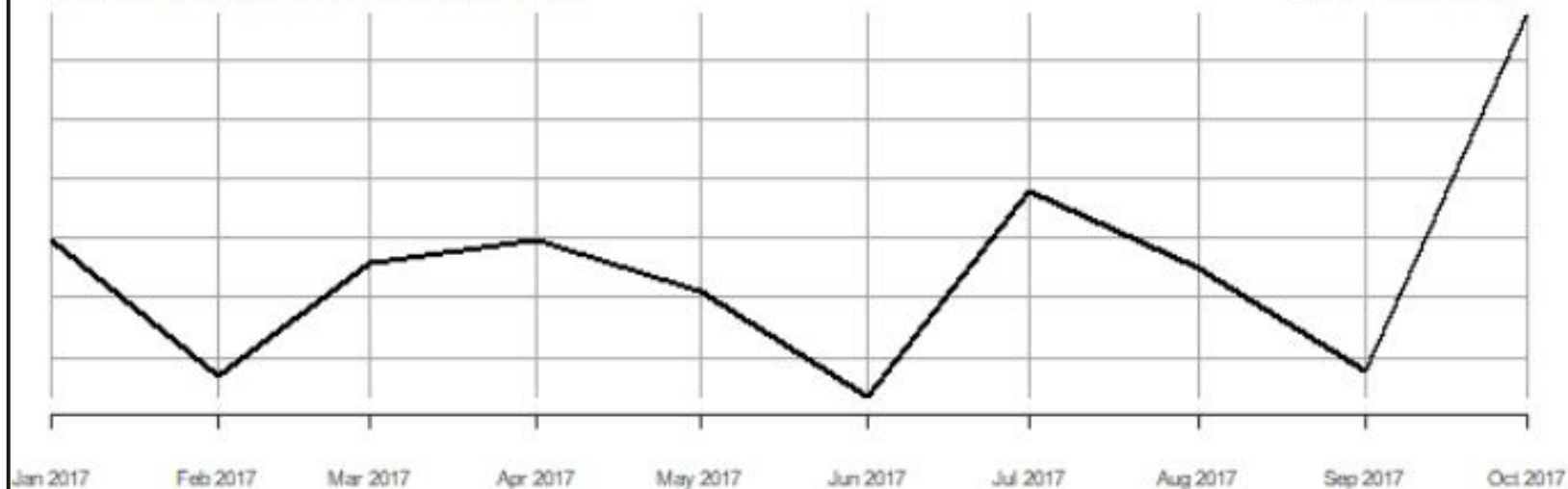


# 2018 FIFA World Cup Russia™

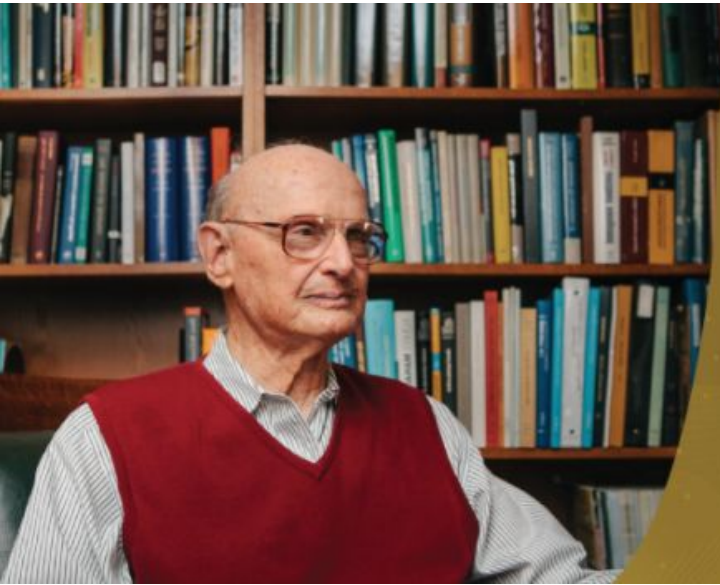
14 June - 15 July

Subperiod Analysis (2017) - Microsoft

Jan 2017 / Oct 2017



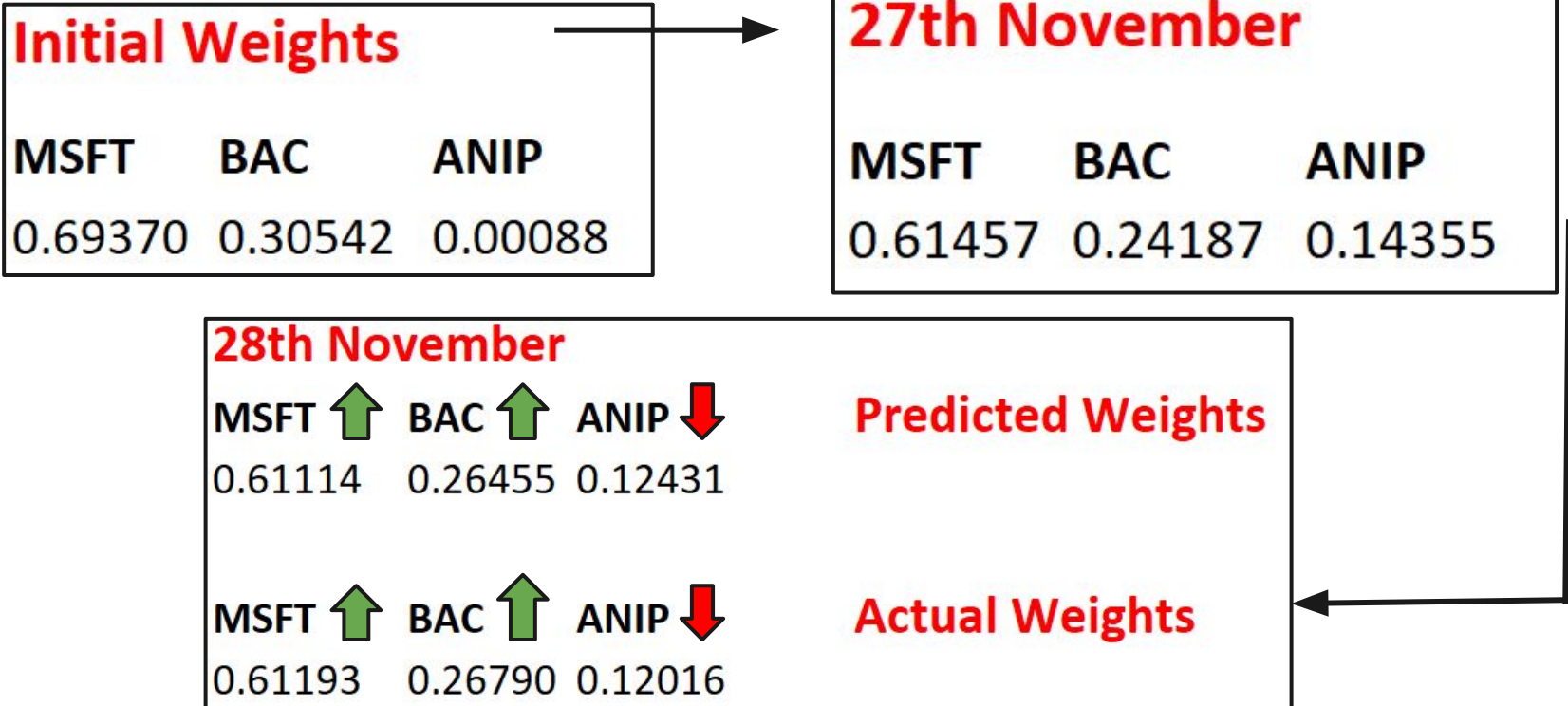
# Modern Portfolio Theory (MPT)



HARRY MARKOWITZ



# Portfolio Analysis | Package: PerformanceAnalytics



# Portfolio Analysis | Package: PerformanceAnalytics



**Variance of Portfolio Returns**

0.00006

**Variance of NASDAQ 100 Returns**

0.00009



# Back Testing :

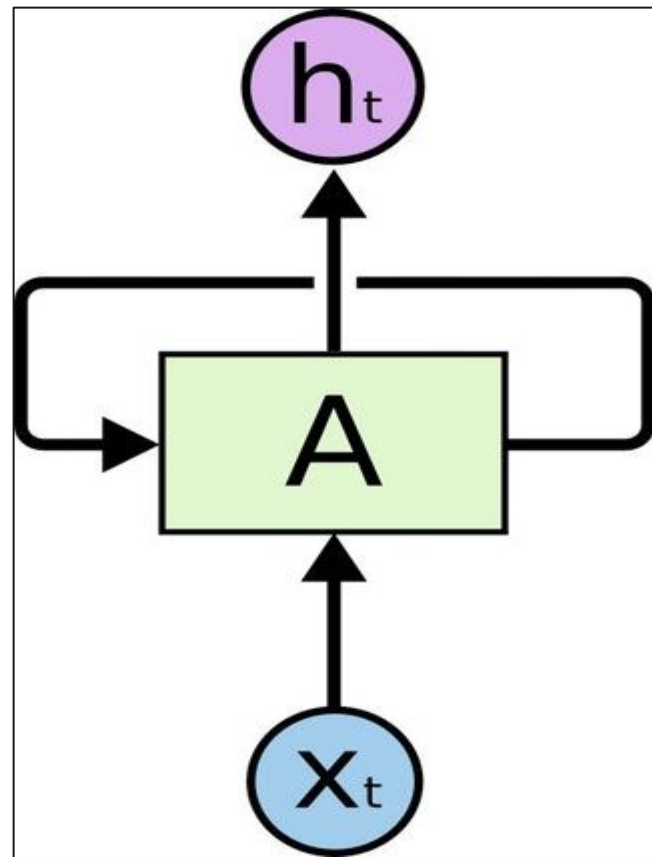
## *Global Financial Crisis*

Name/Date	2008-09-15	2008-09-16	2008-09-17	2008-09-18	2008-09-19
bac_act	24.10000	26.90000	24.5	27.8	34.13
bac_pred	31.00000	24.00000	26.6	24.9	27.7
RMSE	0.80000	3.32000			
anip_actual	129.22000	127.23000	126.6	133.67	147.1
anip_pred	123.59998	129.79501	125.07104	122.78887	127.19079
RMSE	1.76000	8.50000			
msft_actual	21.10000	20.40000	19.3	19.1	19.8
msft_pred	38.68474	38.53389	38.49778	38.34123	38.48040
RMSE	14.76000	15.56000			

“Those who have knowledge, don't predict. Those who predict, don't have knowledge”

--Lao Tzu, 6th Century BC Chinese Poet.

1. **Recurrent Neural Networks : LSTM** (Long Short Term Memory ) has provided results for **sequence dependent data**.<sup>[2]</sup>
2. LSTM other Time Series Applications :
  - a. Question answering.
  - b. Video to text.
  - c. Handwriting generation.
  - d. **Stock Market Prediction.**



# References



1. Asep Juarna, Adi Kuswanto, Mohammad Abdul Mukhyi, and Raden Supriyanto, “Curve Fitting and Stock Price Prediction Using Least Square Method”, ICBLCRS’15
2. Murtaza Roondiwala, Harshal Patel , Shraddha Varma, “Predicting Stock Prices Using LSTM”, IJSR 2017
3. Jana Cipan, “EVALUATING FORECAST ACCURACY”
4. [www.investopedia.com/](http://www.investopedia.com/)
5. <https://mrjbq7.github.io/ta-lib/>
6. [www.schroders.com/](http://www.schroders.com/)
7. [https://en.wikipedia.org/wiki/Time\\_series](https://en.wikipedia.org/wiki/Time_series)
8. <https://www.r-bloggers.com/forecasting-stock-returns-using-arima-model/>
9. <http://www.tradinggeeks.net/2014/07/technical-analysis-with-r/>



*Any Questions ??*