Big Data Applications Symposium - Fall 2017

Project Name: Technical & Fundamental Analysis of S&P 500 companies on NYSE

Team: Aniruddh Khera, Priyanka Vaidya

Abstract: This project examines the volatility of S&P500 companies' stock and carry out both fundamental and technical analysis of the companies listed on the index to analyze the best strategies to increase returns, make portfolios of various companies and predict stock market movement on rolling basis. The ultimate goal is to automate the entire trading process by extending and the scoping to the original plan.

Technical & Fundamental analysis of S&P 500 companies on NYSE

Motivation

Who are the users of this application?

Financial Analysts, Personal Trader, Hedge Funds.

Who will benefit from this application?

Investors, Hedge Funds.

Why is this application important?

- The application is important because it tries to draw a correlation between fundamental data and daily of stock price returns of S&P500
- Through this correlation we can find a process to systematically invest in stocks that give a better return than the benchmark index (S&P 500)

Technical & Fundamental analysis of S&P 500 companies on NYSE

Remediation

What actuation(s) or remediation actions are performed by the application?

- Try to scale the application to incorporate numerous other financial parameters and check which is increasing the portfolio returns for a user.
- Train the machine learning model with the new incoming data. Or better consider a model which required less training.

Technical & Fundamental analysis of S&P 500 companies on NYSE

Data Sources

Name: Price-Adjusted-Split by **Yahoo! Finance**

Description: Contains daily (open, high, low & close) prices with volume traded of S&P500 companies

Size of data: 50 MB

Name: Fundamentals by **NASDAQ**

Description: Contains 78 fields on each S&P500 companies' annual cash ratio, return on equity , taxes filed, etc.

Size of data: 1.09MB

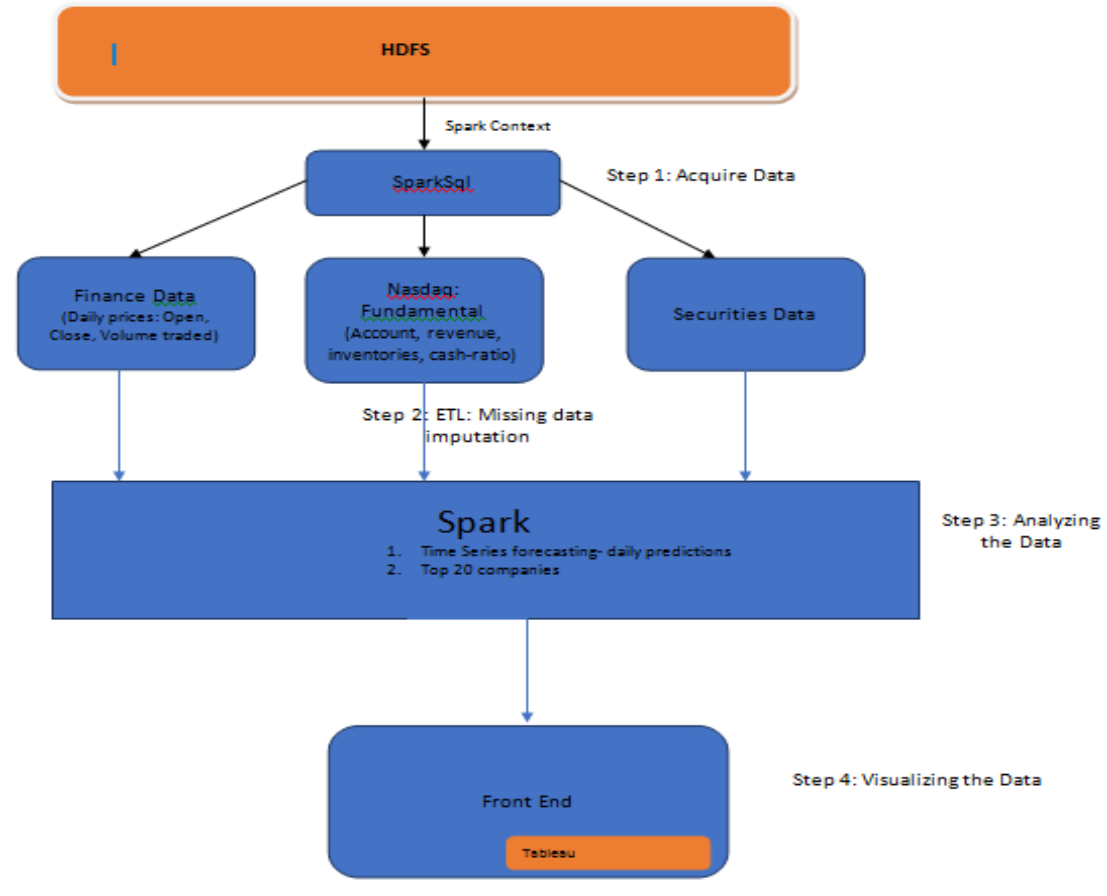
Name: Securities by **EDGAR SEC**

Description: Provides sector and sub sector of each S&P500 company.

Size of data: 51.3 KB

Technical & Fundamental analysis of S&P 500 companies on NYSE

Design Diagram



Platform(s) on which the application runs:

NYU HPC cluster
VM & Standalone to experiment with sbt

Technical & Fundamental analysis of S&P 500 companies on NYSE

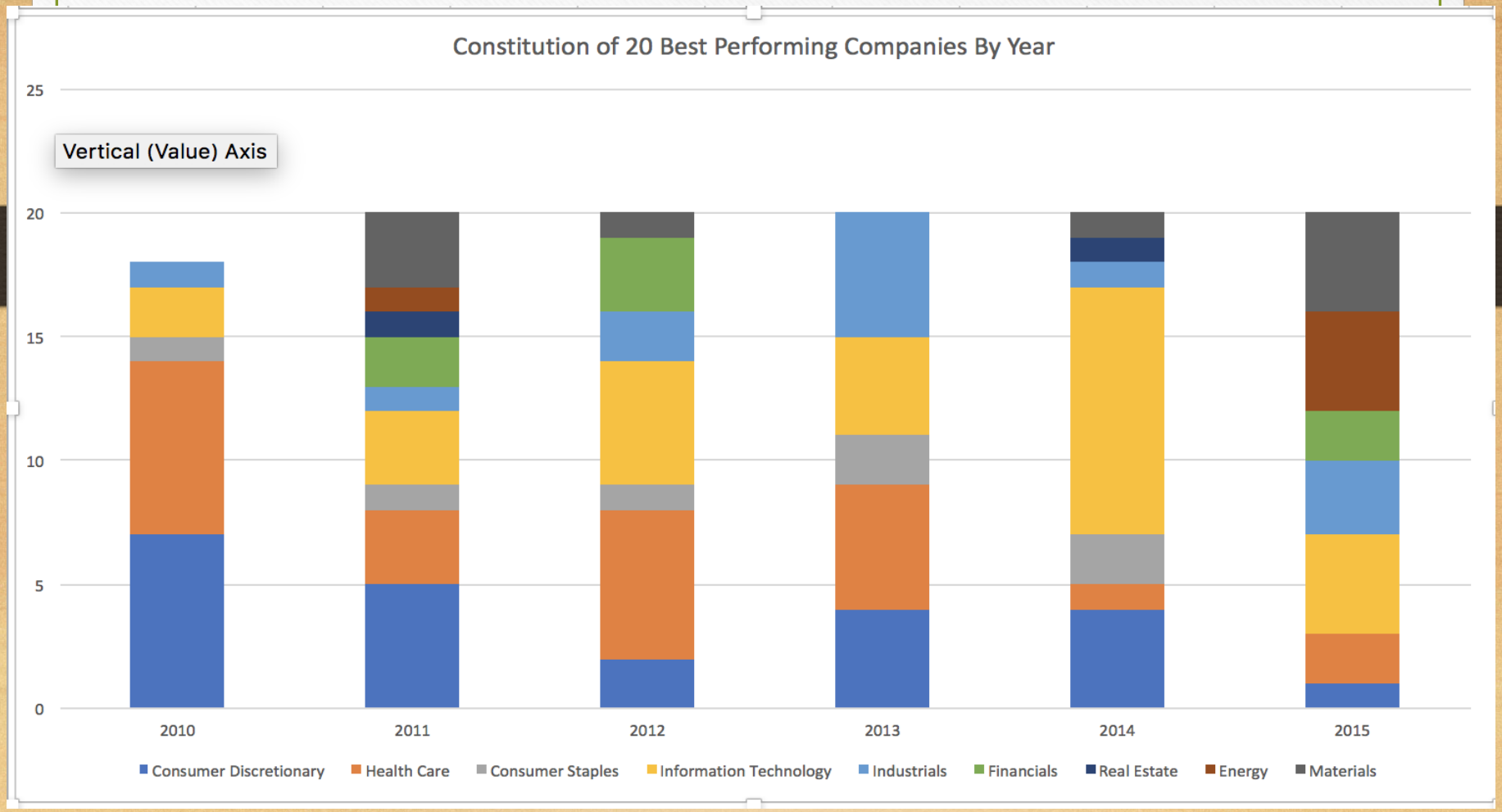
Experiment/Result 1: (Fundamental Analysis)

The number of **Technological** companies in the top 20 performers on S&P500 index (based on calculated value **Return of Investment**) is increasing almost monotonically across years.

$$Return\ of\ Investment\ (R.O.I) = \frac{Current\ yr\ closing\ price}{Prev\ yr\ closing\ price} - 1$$

Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 1: (Fundamental Analysis)



Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 2: (Fundamental Analysis)

We use **Cash Ratio** (C.R.) of the previous year as a predictor of the stock performance. We choose the top N and bottom N companies from the S&P 500 based on **C.R.** to take long and short positions in and rebalanced at the end of every year.

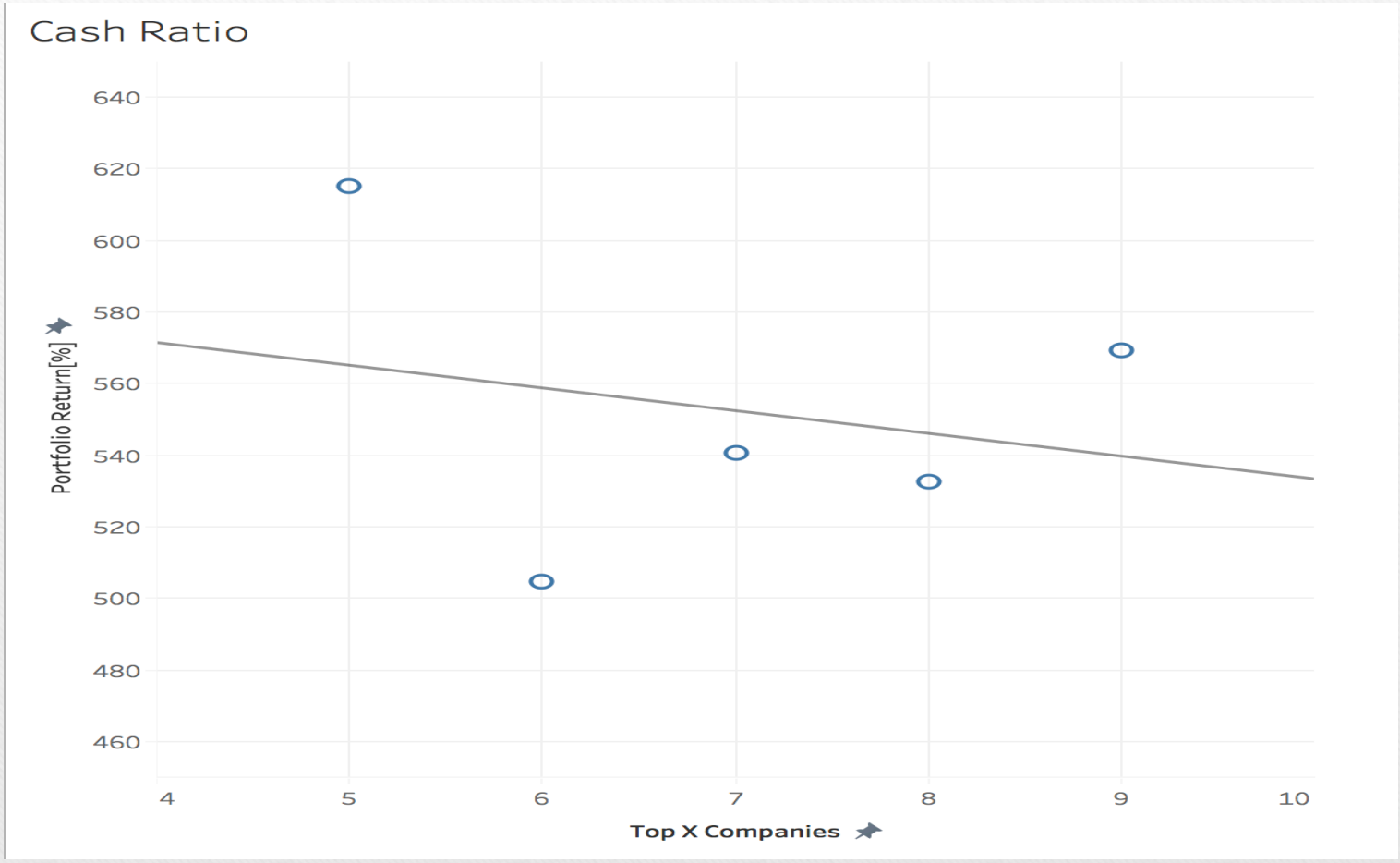
This gave us 2 insights :

1. Cash Ratio is an inverse predictor of next year's returns, which follows the **Mean Reversion** financial theory.

2. The effects of diversification tend to fade after top 5 companies and returns in fact diminishes if we increase the top N companies.

Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 2: (Fundamental Analysis)



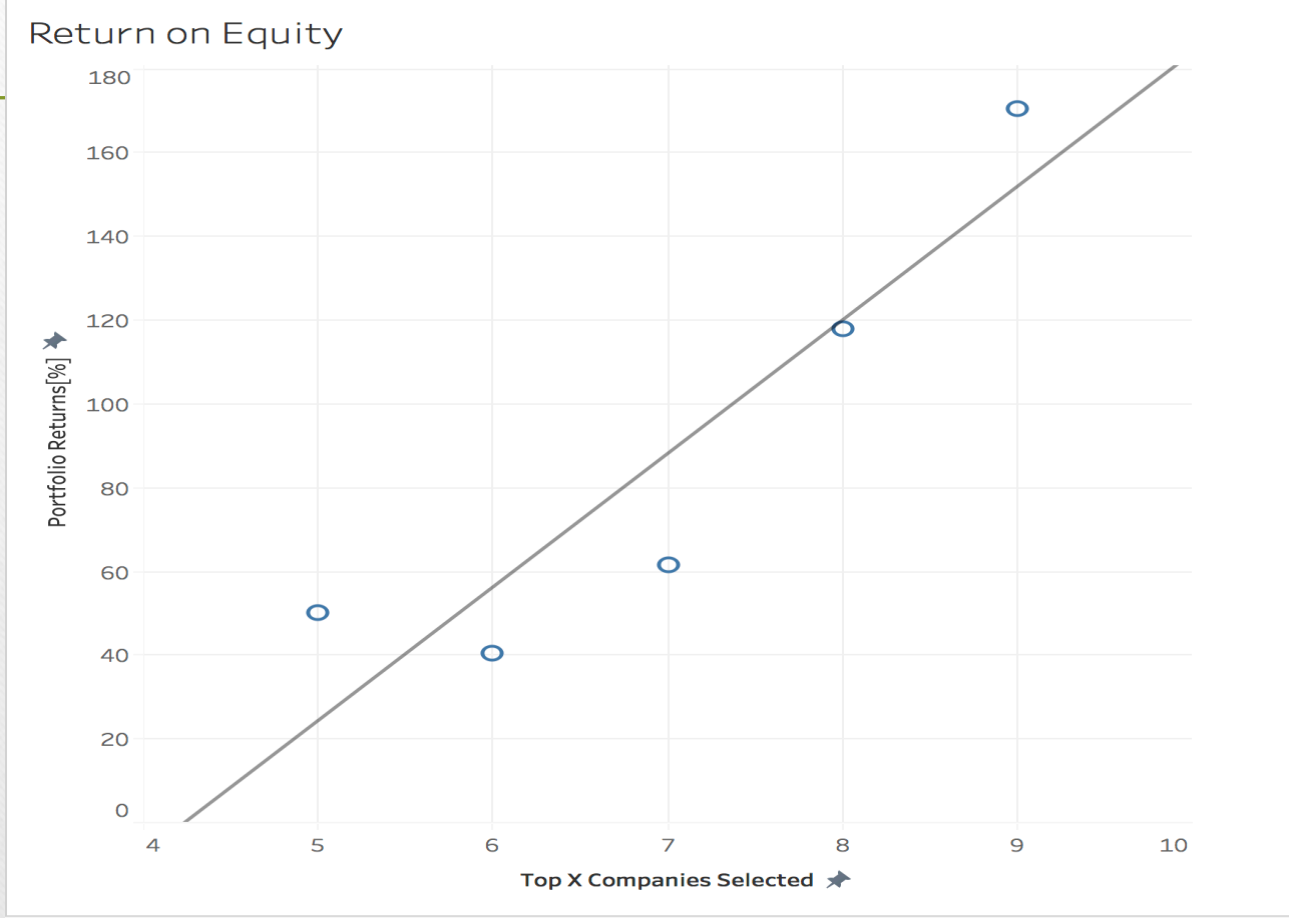
Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 3: (Fundamental Analysis)

ROI showed us that maximum diversification of companies across sector yields better Portfolio Returns.

Technical & Fundamental analysis of S&P 500 companies on NYSE

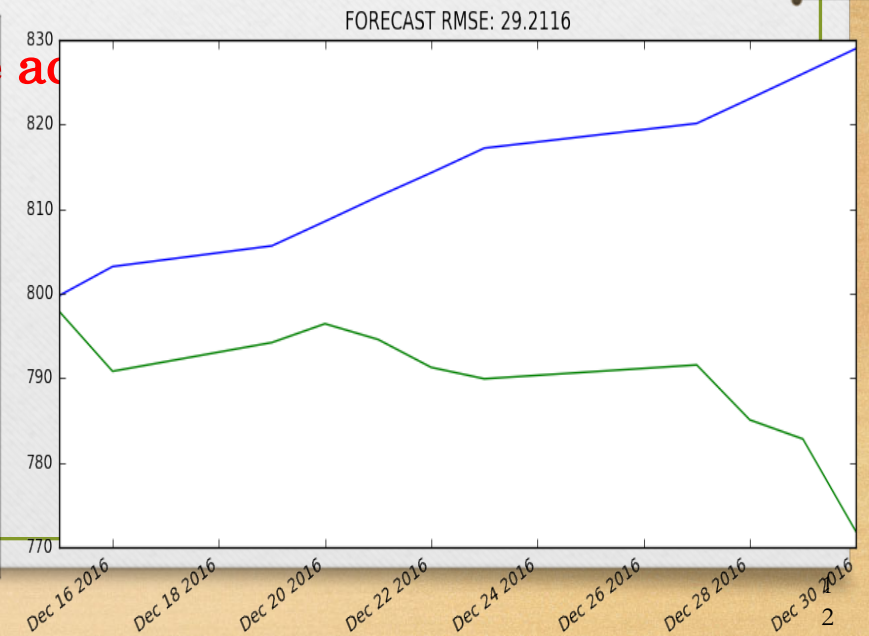
Experiment/Result 3: (Fundamental Analysis)



Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 4: (Technical Analysis)

Using the well known statistical model: **ARIMA (Auto-regressive Integrated Moving Average)** as machine learning model, we were able to forecast next/last15 days stock closing value of S&P companies with a decent yet could be improved accuracy.



Technical & Fundamental analysis of S&P 500 companies on NYSE

Experiment/Result 5:

Timeseries RDD: How do we lay out across machines?

ate	symbol	open	close	low	high	volume
1/5/2016 0:00	WLTW	123.43	125.84	122.31	126.25	2163600
1/6/2016 0:00	WLTW	125.24	119.98	119.94	125.54	2386400
1/7/2016 0:00	WLTW	116.38	114.95	114.93	119.74	2489500
1/8/2016 0:00	WLTW	115.48	116.62	113.5	117.44	2006300
1/11/2016 0:00	WLTW	117.01	114.97	114.09	117.33	1408600
1/12/2016 0:00	WLTW	115.51	115.55	114.5	116.06	1098000
1/13/2016 0:00	WLTW	116.46	112.85	112.59	117.07	949600
1/14/2016 0:00	WLTW	113.51	114.38	110.05	115.03	785300
1/15/2016 0:00	WLTW	113.33	112.53	111.92	114.88	1093700
1/19/2016 0:00	WLTW	113.66	110.38	109.87	115.87	1523500

	5:00 PM	6:00 PM	7:00 PM	8:00 PM	9:00 PM	10:00 PM
GOOG	$523		$524	$600	$574	$400
AAPL	$384	$384	$385	$385	$378	$345
YHOO	$40	$60			$70	$80
MSFT	$134	$138	$175	$178	$123	$184
ORCL	$23	$30	$35	$45	$38	

Helpful generally in any analysis correlation/ lag analysis, machine learning on time series by minimizing/no shuffles across the cluster.

Reference: https://www.youtube.com/watch?v=tKkneWcAIqU

Technical & Fundamental analysis of S&P 500 companies on NYSE

Obstacles

1. Version clashes of dependencies
2. Sbt! Easier to use maven. JDK compatibility

Technical & Fundamental analysis of S&P 500 companies on NYSE

Summary

- With the analytics ran, we found significant correlation between fundamental data and daily stock price returns of S&P500 at the same time we are able to predict the stock movement using machine learning. Therefore we can find a process to systematically invest in stocks that give a better return than the benchmark index (S&P 500).

Acknowledgements

- HPC Dumbo support team
- Spark-ts library owner, Advance Analytics writer Sandy Ryza. Spark summit, 2016

Technical & Fundamental analysis of S&P 500 companies on NYSE

References

[1] Stock Price Prediction Using the ARIMA Model : Ayodele A. Adebiyi., Aderemi O. Adewumi, Charles K. Ayo 2014 [http://ijssst.info/Vol15/No4/data/4923a105.pdf]

---

[2] Estimating stock market volatility using asymmetric GARCH models: Dima Alberga , Haim Shalita, and Rami Yosef 2008 [http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.363.909&rep=rep1&type=pdf]

[3] Short-sellers, fundamental analysis, and stock returns

Patricia Mdechowa, Amy Phutton, LisaMeulbroekb, Richard Gsloana [http://www.sciencedirect.com/science/article/pii/S0304405X01000563/pdfft?md5=27ad785b59f742ca2eaed6eb36e7cfd6&pid=1-s2.0-S0304405X01000563-main.pdf]

[4] Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network: YudongZhang, LenanWu [http://www.sciencedirect.com/science/article/pii/S095741740800852X/pdfft?md5=7a6b3ce461c097996b2cf8cddb46bf6f&pid=1-s2.0-S095741740800852X-main.pdf]

Data source: https://www.kaggle.com/dgawlik/nyse/data

Technical & Fundamental analysis of S&P 500 companies on NYSE

Thank you!