SMARTFOLIO.. A Portfolio Management Tool

Presented by – Ankita Mitkari , Anushree Bagwe , Aparna Sree , Sonam Bhatia (MSIS Big Data Analytics Program)

Georgia State University | J. Mack Robinson College of Business

Introduction

•Trading in a volatile stock market can feel like a gamble to some people due to lack of direction to invest smartly.
•SmartFolio, a Portfolio Management Tool attempts to solve this problem by leveraging machine learning to help traders make right investment decision.
• 10 years' worth Historical price and Company financial data for companies of Dow 30 index is used by SmartFolio.
•SmartFolio provides an interactive dashboard for quick overview of the market along with regression reults from several models like Multi Linear Regression,Gradient Boosted, LSTM and Reinforcement Learning.

Architecture

This project is divided into 3 modules.

1st Module –Fetch raw data from data sources and store into AWS S3 data repository
2nd Module - Prepare raw data from S3 by cleaning, transforming and feature engineering and load back to S3
3rd Module – Pass the prepared data through regression models and store results back to S3. Finally, visualize meaningful insights from raw data and regression results into Tableau.

Data Sources

- This tool uses data from different data sources such as historical stock prices, company's financial performance report and Economic data for democratization of intelligent Asset Management.
- The historical data contains stock value data from 2009-2019 and was extracted using API provided by Alphavantage.
- The financial company data was obtained from Wharton Research Data services and has financial statements from 1986.
- The economic data was pulled from Federal Reserve Economic Data for USA's GDP performance since 1947.

Feature Engineering

SMA 7, 14, 28, 50, 200, Delta = SMA 200 – SMA 50
Date as a feature
Month – capture seasonal variations
Year – relative to start .
Week Seq – Week number from start of week, I.e, week 0 to week 387
- Feature engineering improved the r2 value of multi regression model

	R2			
	No feature engineering	feature engineering	Status	Comments
BA	0.055728179	0.960221141	R2 improved	
JPM	0.578810354	0.937671262	R2 improved	
GS	0.783632057	0.932999284	R2 improved	
PG	0.751004669	0.665153354	R2 decreased	Maybe overfitting

Dashboard

Home Page of SmartFolio

Dow 30 Daily Performance

Company Financial Analysis

Multi Regression

Predict outcome of response variable
• "close" value of each Dow 30 stock
Predict for all Dow 30 companies
- Created 30 different multiregression model

•One model for each of the stocks
- Predicted the close value

Results
- Predicts 7 days into future
- Used extensive feature engineering
- Highest R2 value 0.96
- Lowest R2 value 0.75

close	DailyDate	close_lag7D
208.072	2010-01-21 00:00:00	194.73
197.75	2010-01-22 00:00:00	195.86
203.075	2010-01-25 00:00:00	199.23
205.94	2010-01-26 00:00:00	192.05
207.884	2010-01-27 00:00:00	195.46
199.29	2010-01-28 00:00:00	194.12
192.063	2010-01-29 00:00:00	196.19
194.73	2010-02-01 00:00:00	195.116
195.86	2010-02-02 00:00:00	198.67
199.23	2010-02-03 00:00:00	200.38

Gradient Boosted

Regression model used in many Kaggle solutions
- Created models for 5 stocks.
- Prone to overfitting

Results
- The results of Gradient boosted was only slightly better than linear regression
- Complex model need not always perform better than simple explainable models

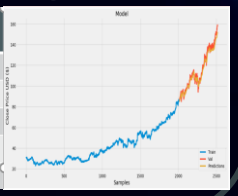
•Required heavy regularization to get decent results

ticker	LR_R2	LR_RMSE
AXP	0.90	2.18
DIS	0.90	2.27
GS	0.93	9.74
JPM	0.94	2.69
BA	0.96	3.65

LSTM

- Trained the close price data for 60 days and predicting the close price value at 61st day.
- The LSTM architecture used for prediction is 4 Layer
- Layer 1 - Number of neurons – 50, Return sequence - True
- Layer 2 - Number of neurons – 50, Return sequence - False
- Layer 3 - Dense with neuron- 25
- Layer 4 – Dense layer with neuron - 1
- Results for Microsoft data show below
- R2 value 0.96

Actual Close	Predictions
85.540001	85.243675
85.949997	85.241837
86.349998	85.276344
87.110001	85.367226
88.190002	85.558388



Reinforcement Learning

A RL agent begins with a fixed sum of money and data on past daily stock prices, over 10 years of time. It will make a buy / sell / hold decision for the long-term maximization of the portfolio value (uninvested cash + present value of stocks). We considered a group of 3 stocks in our portfolio.
- Input : Daily Prices for MSFT , APPL , IBM
- State vector : Number of shares , Price of each share , Uninvested Cash
- Action : Buy , Sell , Hold
- Reward : Change in portfolio value
- Modelling : Q(s`,a`) = R + gamma . Max Q(s,q)
- Output : Distribution of our portfolio value over the number of episodes

References

- https://aws.amazon.com/s3/faqs/
- Xiong et al., Practical Deep Reinforcement Learning Approach for Stock Trading
- https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db