Investment Portfolio Construction, Optimization, and Performance Analysis Using Machine Learning.

Will Bailey, Chris Cline, Alex Cole, Rachel Piraino, Dr. Hum Nath Bhandari

Gabelli School of Business, Department of Computer Science, Department of Mathematics
Roger Williams University, Bristol, RI, USA

RWU SASH 2023

Team of Collaborators

- Research Students: Will Bailey (2024), Chris Cline (2023), Alex Cole (2023), Rachel Piraino (2023).
- Advisor: Dr. Hum Nath Bhandari
- Course: Math 479 - Senior Data Science Capstone

Research Background

- Financial markets are one of the areas most focused on by Machine Learning developers in recent years.
  - Creating algorithms to construct and optimize stock portfolios are at the forefront of AI research.
  - Our research attempts to build a portfolio using modern Machine Learning techniques that is able to outperform our benchmark, the S&P 500 Index.
- The Stock Market is described by many as being "Semi Efficient"
  - This means that with all publicly available information, it's near impossible to outperform the market consistently.
  - The problem that Machine Learning research faces today is how can we create an algorithm that will predict stock price movement and trade accordingly.

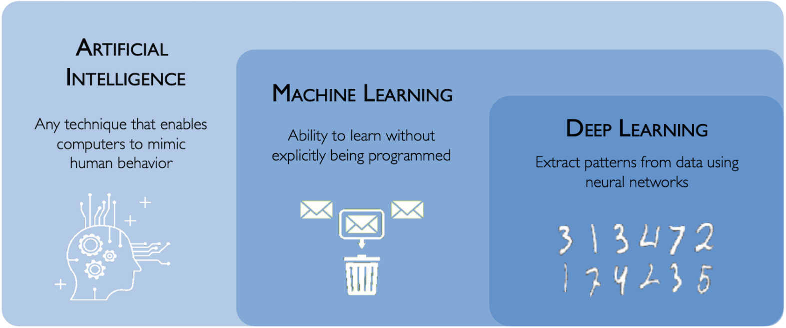
Our Goal

Phase One : Predict the daily closing price of each individual stocks in the S&P 500 Top 50 Index over the time span of 2018-2022.

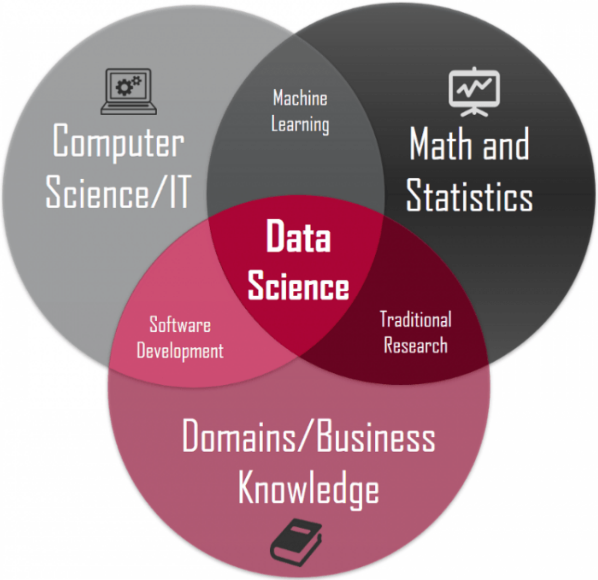
- We used Sentiment, Technical, and Fundamental data for each stock to provide us with a closing price prediction.

Phase Two : Construct a portfolio based on our closing price prediction to maximize returns while minimizing risk.

- We built an efficient portfolio each trading day based on our return prediction and a calculated risk measure.


ARTIFICIAL INTELLIGENCE
Any technique that enables computers to mimic human behavior
MACHINE LEARNING
Ability to learn without explicitly being programmed
DEEP LEARNING
Extract patterns from data using neural networks
3 1 3 4 7 2
1 7 4 2 3 5

Interdisciplinary Field of Study



Machine Learning in the Financial Sector

- Investment monitoring and recommendations

  → Training models to make recommendations based off of analyzing past data.

- Fraud detection

  → Constantly assessing customer behavior, so that when an abnormality occurs it can flag the 'customer'.

- Algorithmic trading

  → AI used in trading by determining patterns, analyzing data, and to generate a model to get the best recommendation.

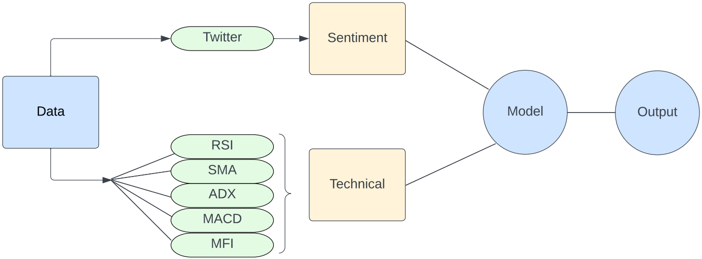
- Pricing, risk/return, and customer decision

  → Similar to our project! Creating models to train with data in hopes to out perform the overall market.

Overall Project Steps

- 1. Collect Sentiment Data
  → Twitter
- 2. Analyze Sentiment Data
  → FINBert sentiment analyzer
- 3. Collect Technical Data
  → SMA, RSI, ADX, MFI, MACD
- 4. Create the model
  → Deep Learning Models: LSTM & GRU
- 5. Training the model
  → Test - Train Split
- 6. Portfolio Evaluation
  → Traditional portfolio performance analysis

Flow Chart


Data
Twitter
Sentiment
RSI
SMA
ADX
MACD
MFI
Technical
Model
Output

Collecting Sentiment Data

- Sentiment of each stock was collected via tweets on Twitter
- To do this we used the SNS scrape library in Python and the Advanced Search function on twitter
  - This function allowed us to only collect tweets with 50 likes or more
- We looked at tweets both from specific companies and tweets of the general public
- Across all 50 stocks we collected 500,000+ tweets

"AAPL $AAPL" min_faves:50 until:2021-12-31 since:201

Top Latest People Photos Videos

6 8 126

TrendSpider @TrendSpider · May 1, 2021

$AAPL #AAPL Nice broadening formation on the sessions chart. Who is looking for a big move from tech? chrt.biz/AAPL/13964sr7y...

13 17 115

Danny Naz | Trading Coach @ThePupOfWallSt · Nov 20, 2021

$AAPL #AAPL Weekly Chart

We had a beautiful bullish marubozu candlestick on the weekly close through our 3.618 Fib extension level. My next big level target is $174.

We also have a MACD crossover w/ a TTM squeeze still setting up.

Charts courtesy of @TrendSpider #PupCharts

Sentiment Analysis

	Date	username	tweet	num_of_likes	num_of_retweet	sentiment
0	2023-03-03 00:30:12+00:00	Microsoft	@addUTILITI 😳 chaos.	0	0	neutral
1	2023-03-03 00:29:16+00:00	Microsoft	@kolide Always.	0	0	neutral
2	2023-03-03 00:28:42+00:00	Microsoft	@FastNFTio 🤝	0	0	neutral
3	2023-03-03 00:27:23+00:00	Microsoft	@Outlook Organization is key. 👏	20	3	positive
4	2023-03-03 00:26:05+00:00	Microsoft	@MicrosoftTeams Of course. 💜	22	1	neutral
5	2023-03-03 00:25:33+00:00	Microsoft	@addUTILITI Answers are needed. 😂	1	0	neutral
6	2023-03-03 00:24:56+00:00	Microsoft	@beingageek 😂	1	0	neutral
7	2023-03-03 00:24:25+00:00	Microsoft	@369volts Good point. 🤔	2	0	positive
8	2023-03-03 00:22:52+00:00	Microsoft	@Dell 👏👏	13	1	neutral
9	2023-03-02 22:00:01+00:00	Microsoft	It's 2023 and we should be normalizing agendas...	415	51	neutral
10	2023-03-02 19:06:08+00:00	Microsoft	When you find out that your coworker pinned yo...	414	35	neutral
11	2023-03-02 18:51:36+00:00	Microsoft	@FastCompany It's an honor to be recognized fo...	1	1	positive
12	2023-03-02 17:30:01+00:00	Microsoft	Custom bookable time in Outlook makes it easie...	129	29	neutral
13	2023-03-02 17:30:00+00:00	Microsoft	POV: you're trying to find available time on y...	188	27	neutral
14	2023-03-02 00:30:41+00:00	Microsoft	@Scorptec Appreciate the honesty. 👏	14	0	positive
15	2023-03-02 00:29:57+00:00	Microsoft	@justinongisrad Not the mental tabs. 😂	31	1	neutral
16	2023-03-02 00:28:43+00:00	Microsoft	@ninja_wiener Sometimes it gets worse before i...	16	0	negative
17	2023-03-02 00:27:29+00:00	Microsoft	@ojemil We love a double digit number.	15	0	positive

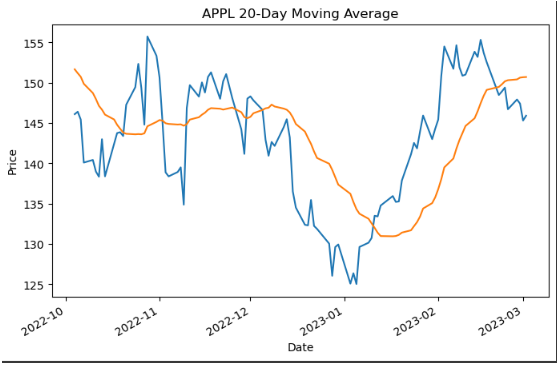
Analyzing Sentiment Data

- Each Tweet was analyzed using FINBERT which is a model pre-trained for financial sentiment analysis
- Each tweet was determined to be negative, neutral, or positive, and assigned a corresponding score (-1, 0, or 1). We then took the average of all sentiment scores for a given day

Date	sentiment_value
2021-12-03	0.007692
2021-12-06	0.052941
2021-12-07	0.103175
2021-12-08	0.160000
2021-12-09	0.008547
2021-12-10	0.106796
2021-12-13	0.104478
2021-12-14	0.029630
2021-12-15	0.061856
2021-12-16	0.016260
2021-12-17	0.084746
2021-12-20	0.066667

Collecting Technical Data

- Used yFinance and PyPI libraries to collect technical data on our stocks
- Calculated our own indicators and process the data to be used as inputs for our model



Model Features Table

Table 2: **Model Features**

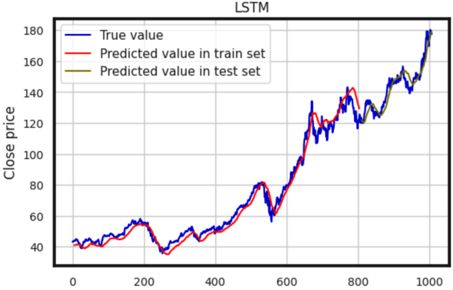
Data	Source	Library
Technical Indicator		
20 Day Simple Moving Average	Yahoo	stock-indicators
Moving Average Convergence Divergence	Yahoo	stock-indicators
Stochastic	Yahoo	stock-indicators
Relative Strength Index	Yahoo	stock-indicators
Average Directional Index	Yahoo	stock-indicators
Money Flow Index	Yahoo	stock-indicators
Average True Range	...	...
Sentiment		
Investor Sentiment	Twitter	Snscrape

LSTM and GRU Architectures

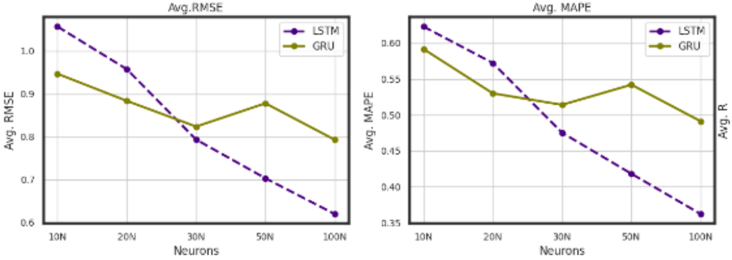
- LSTM
  - An efficient ML/DL technique for sequential data prediction.
  - Improved version of Recurrent Neural Networks(RNN).
  - Uses memory cells to retain the information for a longer time.
  - Computation in four gates: input gate, forget gate, change gate, and output gate
- GRU
  - A simplified version of the LSTM.
  - Three gates in GRU: update gate, reset gate, and change gate.
  - The update gate is equivalent to the forget gate and input gate of LSTM. It is responsible for the long-term dependencies.
  - The reset gate is responsible for the short-term dependencies.

Implementation of the Model

- Implemented both GRU and LSTM and ran with various number of neurons for training purposes
- 80% of original data is training and other 20% is test data to compare to the prediction made.
  - LSTM Model graph for Apple stock below:



Model Selection



Model Evaluation

- LSTM
  - A 100 Neuron LSTM provided us with the best results.
  - RSME was used as our selection criteria, with the lowest RMSE model chosen.
- Adam optimizer: works best with sparse data sets while not needing to tune the learning rate

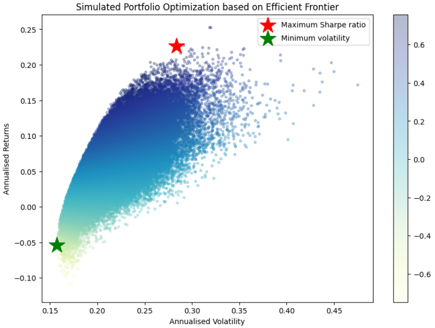
RMSE	MAPE	R
.62	.36	.99

Portfolio Construction Steps

- Process
  - Asset Selection
    - S&P 500 Top 50 Index stocks
    - Benchmark: S&P 500 Index
  - Investment Guidelines
    - Long only strategy
  - Portfolio monitoring and re-balancing
    - Model return prediction used to optimize an efficient portfolio daily maximizing Sharpe ratio
    - Performance evaluation at the end of our time horizon

Portfolio Construction

- Linear programming to construct a portfolio with optimized sharpe ratio daily
- Each day 10,000 random portfolios were generated, and the one with the best sharpe ratio was selected for that day's weighting scheme

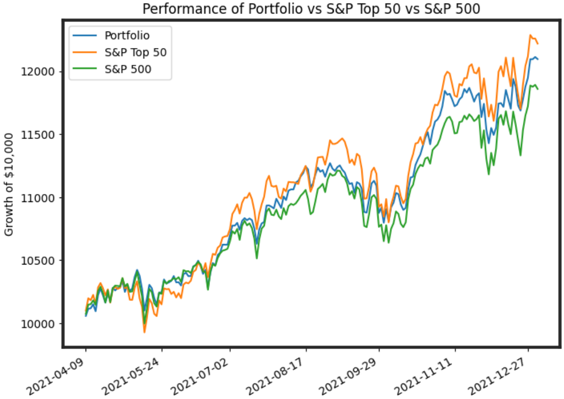


Portfolio Performance

Measure	Portfolio	S&P Top 50	S&P 500
Holding Period Return	20.95%	22.18%	18.58%
Annualized Return	29.38%	31.17%	25.96%
Beta	0.83	1.01	1.00
Std Dev.	10.65%	12.65%	12.03%
Holding Period Alpha	5.03%	3.52%	-%
Annualized Alpha	7.15%	5.09%	-%
Holding Period Treynor Ratio	0.21	0.22	0.19
Annualized Treynor Ratio	0.31	0.31	0.26
Sharpe Ratio	2.43	2.19	1.87

Portfolio and Benchmarks Line Chart

- Portfolio and Benchmark Raw Return 4/1/21 - 12/31/21
  - Portfolio: 20.95%
  - S&P Top 50: 22.18%
  - S&P 500: 18.58%



Reference 1:

Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R.Dahal, Rajendra K.C.Khatri. Predicting Stock Market Index using LSTM, Machine Learning with Applications, Elsevier, 2022.

Machine Learning with Applications 9 (2022) 100320

Contents lists available at ScienceDirect

ELSEVIER

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Predicting stock market index using LSTM

Hum Nath Bhandari[a], Binod Rimal[b,*], Nawa Raj Pokhrel[c], Ramchandra Rimal[d], Keshab R. Dahal[e], Rajendra K.C. Khatri[f]

[a] Department of Mathematics, Roger Williams University, Bristol, RI, USA
[b] Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA
[c] Department of Physics and Computer Science, Xavier University of Louisiana, New Orleans, LA, USA
[d] Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, TN, USA
[e] Department of Statistics, Truman State University, Kirksville, MO, USA
[f] Department of Applied Mathematics, Philander Smith College, Little Rock, AR, USA

ARTICLE INFO

Keywords:
Stock market index
LSTM
Prediction
Machine learning
Deep learning
Denoising

ABSTRACT

The rapid advancement in artificial intelligence and machine learning techniques, availability of large-scale data, and increased computational capabilities of the machine opens the door to develop sophisticated methods in predicting stock price. In the meantime, easy access to investment opportunities has made the stock market more complex and volatile than ever. The world is looking for an accurate and reliable predictive model which can capture the market's highly volatile and nonlinear behavior in a holistic framework. This study uses a long short-term memory (LSTM), a particular neural network architecture, to predict the next-day closing price of the S&P 500 index. A well-balanced combination of nine predictors is carefully constructed under the umbrella of the fundamental market data, macroeconomic data, and technical indicators to capture the behavior of the stock market in a broader sense. Single layer and multilayer LSTM models are developed using the chosen input variables, and their performances are compared using standard assessment metrics–Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Correlation Coefficient (R). The experimental results show that the single layer LSTM model provides a superior fit and high prediction accuracy compared to multilayer LSTM models.

Reference 2:

Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R.Dahal. LSTM-SDM: An integrated framework of LSTM implementation for sequential data modeling, 2022.

Software Impacts 14 (2022) 100396

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

ELSEVIER

SOFTWARE IMPACTS

Original software publication

LSTM-SDM: An integrated framework of LSTM implementation for sequential data modeling

Check for updates

Hum Nath Bhandari [a], Binod Rimal [b,*], Nawa Raj Pokhrel [c], Ramchandra Rimal [d], Keshab R. Dahal [e]

[a] *Department of Mathematics, Roger Williams University, Bristol, RI, USA*
[b] *Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA*
[c] *Department of Physics and Computer Science, Xavier University of Louisiana, New Orleans, LA, USA*
[d] *Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, TN, USA*
[e] *Department of Statistics, Truman State University, Kirksville, MO, USA*

ARTICLE INFO

Keywords:
LSTM
Prediction
Deep learning
Sequential data modeling
Time series
Data science

ABSTRACT

LSTM-SDM is a python-based integrated computational framework built on the top of Tensorflow/Keras and written in the Jupyter notebook. It provides several object-oriented functionalities for implementing single layer and multilayer LSTM models for sequential data modeling and time series forecasting. Multiple subroutines are blended to create a conducive user-friendly environment that facilitates data exploration and visualization, normalization and input preparation, hyperparameter tuning, performance evaluations, visualization of results, and statistical analysis. We utilized the LSTM-SDM framework in predicting the stock market index and observed impressive results. The framework can be generalized to solve several other real-world time series problems.

Conclusion

- Our model was able to predict stock returns with enough accuracy to significantly reduce risk while preserving returns
- The portfolio we constructed was able to outperform our benchmark on a raw and risk adjusted basis
- We will soon apply new optimization techniques to our model in an attempt to improve the accuracy of our predictions
- Significant research has shown that Machine Learning models are able to predict stock performance to a degree where the implementation of such prediction can improve the optimization of any stock portfolio

Thank you

Any Question?
Thank you for listening!

Special shout out to Dr. Bhandari for his guidance with this project over this past semester

References

1. **Hum Nath Bhandari**, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R.Dahal, Rajendra K.C.Khatri. Predicting Stock Market Index using LSTM, Machine Learning with Applications, Elsevier, 2022.
2. **Hum Nath Bhandari**, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R.Dahal. LSTM-SDM: An integrated framework of LSTM implementation for sequential data modeling, 2022.
3. Source of Some Images: Google Images.