BULL AND BEAR

A SYSTEM FOR COMPARATIVE ANALYSIS OF DIFFERENT STOCK PREDICTION METHODOLOGIES FOR BEGINNER INVESTORS

Team: Ishan Jain, Tanmay Goel, Mohibul Hassan, Nimisha Goyal
Advised by - Professor David Rossiter
WHAT IS INVESTING?

Investing is the act of committing money or capital to an endeavour (a business, project, real estate, etc.), with the expectation of obtaining an additional income or profit.

- Investopedia

MOST POPULAR INVESTING PRACTICES

BONDS  GOLD  STOCKS  REAL ESTATE  MUTUAL FUNDS
INVESTING IN STOCKS IS A GOOD IDEA!

S&P 500 Index
INDEXSP: .INX

2,926.17 -1.08 (0.037%) 🔻
25 Apr, 4:50 pm GMT-4 - Disclaimer

1 day 5 days 1 month 6 months YTD 1 year 5 years Max

S&P 500 rose 300% since the 2008-09 financial crisis.
**BARRIERS OF ENTRY FOR BEGINNERS**

1. Overwhelming and confusing information

2. Paid services (websites, wealth managers etc.)

3. Predictions are not verified with data

4. Low level understanding of analyzing news and public sentiment
Bull and Bear

Compare effectiveness of prediction techniques

Educate beginner users

Easy to use and trustworthy
1. Providing predictions made by different indicators within Technical Analysis

- EMA
- RSI
- Bollinger Bands
- MACD
OUR OBJECTIVES

2. Performing Sentiment Analysis on financial news and tweets related to the stocks
3. Comparing these techniques with each other and informing the user about the effectiveness of each prediction technique over a period of time
OUR OBJECTIVES

4. Encompassing all the above features in an aesthetically pleasing and intuitive user interface
TECHNICAL ANALYSIS

Financial analysis that uses patterns in market data to identify trends and make predictions

TYPES OF TECHNICAL INDICATORS

- Trend
- Momentum
- Volatility
- Volume
Choosing the Technical Indicators

Criteria Used

- Variety
- Popularity
- Effectiveness

Bollinger Bands
Exponential Moving Average
Relative Strength Index
Moving Average Convergence Divergence
Backtesting

Aim: To find the parameters that give the highest returns for each indicator

For Each Indicator

- Data for past 13 years for the 30 stocks in Dow Jones
- Pre-determined range of parameters
- Find returns for each parameter
- Store to an Excel file

<table>
<thead>
<tr>
<th>period</th>
<th>deviation</th>
<th>gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.1</td>
<td>215462.718</td>
</tr>
<tr>
<td>5</td>
<td>1.2</td>
<td>205877.898</td>
</tr>
<tr>
<td>5</td>
<td>1.3</td>
<td>205941.493</td>
</tr>
<tr>
<td>5</td>
<td>1.4</td>
<td>193697</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>206536.207</td>
</tr>
<tr>
<td>5</td>
<td>1.6</td>
<td>177359.596</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>167700.662</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>142840.485</td>
</tr>
<tr>
<td>5</td>
<td>1.9</td>
<td>109996.559</td>
</tr>
</tbody>
</table>
BACKTESTING

Aim: To find the parameters that give the highest returns for each indicator

FOR EACH INDICATOR

Take the top 10% returns → Calculate median → Use this best parameter in the web app

<table>
<thead>
<tr>
<th>period</th>
<th>deviation</th>
<th>gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.1</td>
<td>215462.718</td>
</tr>
<tr>
<td>5</td>
<td>1.2</td>
<td>205877.898</td>
</tr>
<tr>
<td>5</td>
<td>1.3</td>
<td>205941.493</td>
</tr>
<tr>
<td>5</td>
<td>1.4</td>
<td>193697</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>206536.207</td>
</tr>
<tr>
<td>5</td>
<td>1.6</td>
<td>177359.596</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>167700.662</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td>142840.485</td>
</tr>
<tr>
<td>5</td>
<td>1.9</td>
<td>109996.559</td>
</tr>
</tbody>
</table>
Bollinger Bands

Consists of an SMA line and two lines plotted 2 standard deviations (+ and -) away

**BUY**
- Close line crosses below the Lower Band

**SELL**
- Close line crosses over the Upper Band
PARAMETER SELECTION FOR BOLLINGER BANDS

VARIABLES

1. Time Period
2. Standard Deviation

RESULT

Best Parameter
14 days, 1.7 SD
Return per year 8.39%

VS

Standard Parameter
20 days, 2 SD
Return per year 7.07%
EXponential MOVing AVERAGE (EMA)

A moving average that places greater weight on the most recent data points

**STRATEGY**

**BUY**
EMA line crosses over the Close line from above

**SELL**
EMA line crosses over the Close line from below
PARAMETER SELECTION FOR EMA

VARIABLES

1. Time Period

RESULT

Best parameter
128 days
Return per year
10.48%

VS

Standard parameter
20 days
Return per year
6.17%
Relative Strength Index (RSI)

- Momentum indicator
- Ranges between 0-100
- Measures the magnitude of recent price changes to analyse whether the stock is overbought or oversold

**Strategy**

**BUY**
RSI crosses the underbought line

**SELL**
RSI crosses the overbought line

An example of RSI graph for Apple stock with standard parameters
# Parameter Selection for RSI

## Variables

1. Time Period
2. Upper Band
3. Lower Band

## Result

**Best parameters**

- 15 days, 78 Upper Band, 37 Lower Band
- Return per year: 9.69%

**VS**

**Standard parameters**

- 14 days, 70 Upper Band, 30 Lower Band
- Return per year: 6.73%
MOVING AVERAGE CONVERGENCE DIVERGENCE (MACD)

- Trend following indicator
- The MACD shows the relationship between two moving averages
- MACD = 12 period EMA − 26 period EMA
- Signal = 9 Day EMA of MACD

**Strategy**

- **BUY**: MACD line crosses above the signal line
- **SELL**: MACD line crosses below the signal line

*MACD graph for Apple stock for the past year with standard parameters*
PARAMETER SELECTION FOR MACD

VARIABLES
1. 1st EMA Period
2. 2nd EMA Period

RESULT
Best parameters
1st EMA Period = 18
2nd EMA Period = 4
Return per year 7.94%

VS
Standard parameters
1st EMA Period = 12
2nd EMA Period = 26
Return per year 4.82%
SENTIMENT ANALYSIS:
FINANCIAL NEWS - WHAT DOES IT MEAN?

For bull market doubters, this relatively safe dividend strategy has provided big returns
7 Hours Ago
For investors who believe the rally is just the decadelong bull run's last hurrah, here's how to play safe when the market gets choppy again.

Hot first-quarter economic growth number challenges Fed as it meets next week
11 Hours Ago
Hotter-than-expected first-quarter growth is likely to revive the debate within the Fed about the direction of policy.

Kudlow says White House will be ‘very aggressive’ in China trade talks with US economy doing so well
14 Hours Ago
National Economic Council Director Larry Kudlow said Friday that the strength of the U.S. economy gives it leverage over China as the two countries try to iron out a trade deal.
SENTIMENT ANALYSIS – A NEW AVENUE TO PREDICT STOCK PRICES?

Kylie Jenner
@KylieJenner

sooo does anyone else not open Snapchat anymore? Or is it just me... ugh this is so sad.

360K 5:50 AM - Feb 22, 2018

75.2K people are talking about this

- $1.3B

Donald J. Trump
@realDonaldTrump

The F-35 program and cost is out of control. Billions of dollars can and will be saved on military (and other) purchases after January 20th.

53.8K 9:26 PM - Dec 12, 2016

22.6K people are talking about this

- $5.8B
SENTIMENT ANALYSIS WORKFLOW

Twitter

IEX API

Tweepy API + Open Source Libraries

Data Collection

Data Selection Algorithms

Data Processing

VADER Sentiment

Sentiment Analysis Scores

Graphical Output

Stock Price Analysis
Making predictions based on market sentiment derived from tweets

1. Collect tweets for all stocks
2. Process the data
3. Quantify the sentiment for each tweet to a value in the range -1 and 1
4. Calculate the average sentiment for each day
5. Compare today’s sentiment to the Buy/ Sell thresholds for that stock
6. Make a suggestion based on the comparison
SENTIMENT ANALYSIS (TWITTER)

Representation of Twitter Sentiment and Stock Value for IBM
Between 04/2018 and 04/2019

Sentiment Analysis (Twitter)

Strategy

Average Daily Sentiment is Greater than a Certain Threshold

BUY

Average Daily Sentiment is Smaller than a Certain Threshold

SELL

IBM Buy Threshold

0.46

IBM Sell Threshold

-0.02
PARAMETER SELECTION FOR TWITTER

VARIABLES
1. Buy Threshold
2. Sell Threshold

RANGE DETERMINATION
- Average Tweet Sentiment = 0.0682
- Standard Deviation = 0.1675
- Positive Range = [0.1, 0.5]
- Negative Range = [-0.22, 0.2]
  (with Increments of 0.04)

RESULT
- Best threshold
- Individual Return per year
  8.49%
COMPARATIVE ANALYSIS

To compare the effectiveness of indicators with each other for every stock in Dow Jones Index:

30 Stocks in Dow Jones

Calculate returns given by EMA, RSI, MACD, Bollinger Bands, and Twitter

Compare the returns and rank the indicators in descending order of returns

Ranking of indicators
Comparative analysis chart for **Boeing**:

- Bollinger Bands perform the best
- Marginal difference among top 3
- Twitter analysis performed the worst
Comparative analysis chart for Goldman Sachs:

- MACD perform the best
- Marginal difference between the top 2
- RSI gives negative return
Comparative analysis chart for Microsoft:

- Twitter analysis perform the best, way better than all the technical indicators
- Bollinger bands performed the worst
Enter the name of the listed company

Submit

BULL AND BEAR IS A COMPARATIVE STOCK PREDICTION PLATFORM
COMPARE DIFFERENT STOCK PREDICTION TECHNIQUES AND DISCOVER THE MOST RELIABLE ONES.

SCROLL DOWN FOR THE LATEST NEWS
**USER TESTING AND RESULTS**

45% Respondents had never invested in the stock market

77% Respondents were not confident in making investment decisions

Majority response towards design - **GREAT** - Majority response towards ease of use

---

**Did you increase your understanding of stock market concepts by using Bull and Bear?**

- Yes: 88.9%
- No: 11.1%

9 responses
KEY TAKEAWAYS

- Parameters for indicators can be customized to improve returns
- Effectiveness of each indicator is different for different stocks
- Sentiment analysis proves to be a new and exciting way to predict prices in the future
- Simple design interface makes the user more confident in understanding difficult concepts
Bull and Bear

Compare effectiveness of prediction techniques

Educate beginner users

Easy to use and trustworthy
A SYSTEM FOR COMPARATIVE ANALYSIS OF DIFFERENT STOCK PREDICTION METHODOLOGIES FOR BEGINNER INVESTORS

Team: Ishan Jain, Tanmay Goel, Mohibul Hassan, Nimisha Goyal
Advised by - Professor David Rossiter
BULL AND BEAR

APPENDIX
## EMA Results Backtest Results

<table>
<thead>
<tr>
<th>Period of Time</th>
<th>Average Return over 13 years</th>
<th>Percentage Return Per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 days</td>
<td>$117,889.98</td>
<td>6.17%</td>
</tr>
<tr>
<td>50 days</td>
<td>$157,707.95</td>
<td>7.55%</td>
</tr>
<tr>
<td>100 days</td>
<td>$209,758.22</td>
<td>9.09%</td>
</tr>
</tbody>
</table>
# Bollinger Bands Backtest Results

<table>
<thead>
<tr>
<th>Period of time</th>
<th>Standard Deviation</th>
<th>Average return over 13 years</th>
<th>Percentage return per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 days</td>
<td>2</td>
<td>$143,155.40</td>
<td>7.07%</td>
</tr>
<tr>
<td>50 days</td>
<td>2</td>
<td>$104,196.06</td>
<td>5.65%</td>
</tr>
</tbody>
</table>
RSI CALCULATION

- \( RSI = 100 - \frac{100}{1 + RS} \)
- \( RS = \text{Relative Strength} = \frac{\text{AvgU}}{\text{AvgD}} \)
- \( \text{AvgU} = \text{average of all up moves in the last N price bars} \)
- \( \text{AvgD} = \text{average of all down moves in the last N price bars} \)
- \( N = \text{the period of RSI} \)
SENTIMENT ANALYSIS TWITTER - HYPOTHESIS

An Analysis Method that uses Backtested Parameters in order to Maximise Gains

<table>
<thead>
<tr>
<th>Stock</th>
<th>Correlation 1</th>
<th>Correlation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>-0.0319</td>
<td>-0.0593</td>
</tr>
<tr>
<td>BA</td>
<td>0.0057</td>
<td>-0.0156</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.0343</td>
<td>0.0016</td>
</tr>
<tr>
<td>GS</td>
<td>-0.0411</td>
<td>0.0389</td>
</tr>
<tr>
<td>CSCO</td>
<td>0.0187</td>
<td>0.0326</td>
</tr>
<tr>
<td>CAT</td>
<td>-0.0155</td>
<td>-0.0151</td>
</tr>
<tr>
<td>CVX</td>
<td>-0.0164</td>
<td>-0.032</td>
</tr>
<tr>
<td>DOW</td>
<td>-0.0451</td>
<td>0.0538</td>
</tr>
<tr>
<td>XOM</td>
<td>0.0402</td>
<td>-0.0045</td>
</tr>
<tr>
<td>HD</td>
<td>-0.0035</td>
<td>0.0017</td>
</tr>
<tr>
<td>INTC</td>
<td>0.0082</td>
<td>-0.0093</td>
</tr>
<tr>
<td>JPM</td>
<td>-0.0135</td>
<td>-0.02055</td>
</tr>
<tr>
<td>MCD</td>
<td>0.0383</td>
<td>0.0099</td>
</tr>
<tr>
<td>MRK</td>
<td>-0.0427</td>
<td>0.01524</td>
</tr>
<tr>
<td>PFE</td>
<td>0.0053</td>
<td>0.0518</td>
</tr>
<tr>
<td>PG</td>
<td>0.0272</td>
<td>-0.0014</td>
</tr>
<tr>
<td>TRV</td>
<td>0.0302</td>
<td>-0.0304</td>
</tr>
<tr>
<td>UNH</td>
<td>0.0427</td>
<td>-0.0047</td>
</tr>
<tr>
<td>WBA</td>
<td>-0.0156</td>
<td>0.0246</td>
</tr>
<tr>
<td>AVE</td>
<td>0.001342105263</td>
<td>0.001962631579</td>
</tr>
</tbody>
</table>

**STRATEGY**

\[
\text{Suggestion}_{\text{today}} = (\text{Sentiment}_{\text{yesterday}} - \text{Sentiment}_{\text{today}})
\]

**REVISED STRATEGY**

Suggestion Based on Daily Average Sentiment
# STOCK EXCLUSION FOR TWITTER

<table>
<thead>
<tr>
<th>Excluded Stock</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCA COLA</td>
<td>- Irrelevant Tweets</td>
</tr>
<tr>
<td></td>
<td>- Extremely Large Dataset</td>
</tr>
<tr>
<td></td>
<td>- Slows Down Program</td>
</tr>
<tr>
<td>NIKE</td>
<td></td>
</tr>
<tr>
<td>APPLE</td>
<td></td>
</tr>
<tr>
<td>WALMART</td>
<td></td>
</tr>
<tr>
<td>VERIZON</td>
<td></td>
</tr>
<tr>
<td>AMERICAN EXPRESS</td>
<td>Less than 10K tweets, not enough to make qualitative analysis</td>
</tr>
<tr>
<td>UNITED TECH</td>
<td></td>
</tr>
<tr>
<td>JNJ</td>
<td></td>
</tr>
<tr>
<td>3M</td>
<td>Confused with ‘3 Million’</td>
</tr>
<tr>
<td>VISA</td>
<td>Confused with ‘Travel Visa’</td>
</tr>
</tbody>
</table>
CHALLENGES FACED

- Finding suitable tweets for sentiment analysis
- Finding enough news articles for sentiment analysis
- Realistic back testing of technical indicators
LIMITATIONS

- Transaction charges ignored while backtesting the indicators
- Past performance does not guarantee future returns
- Sentiment Analysis as of today is unable to detect sarcasm, idioms etc.
- Tweets for sentiment analysis are not completely reliable
COMPARISON WITH OTHER APPS