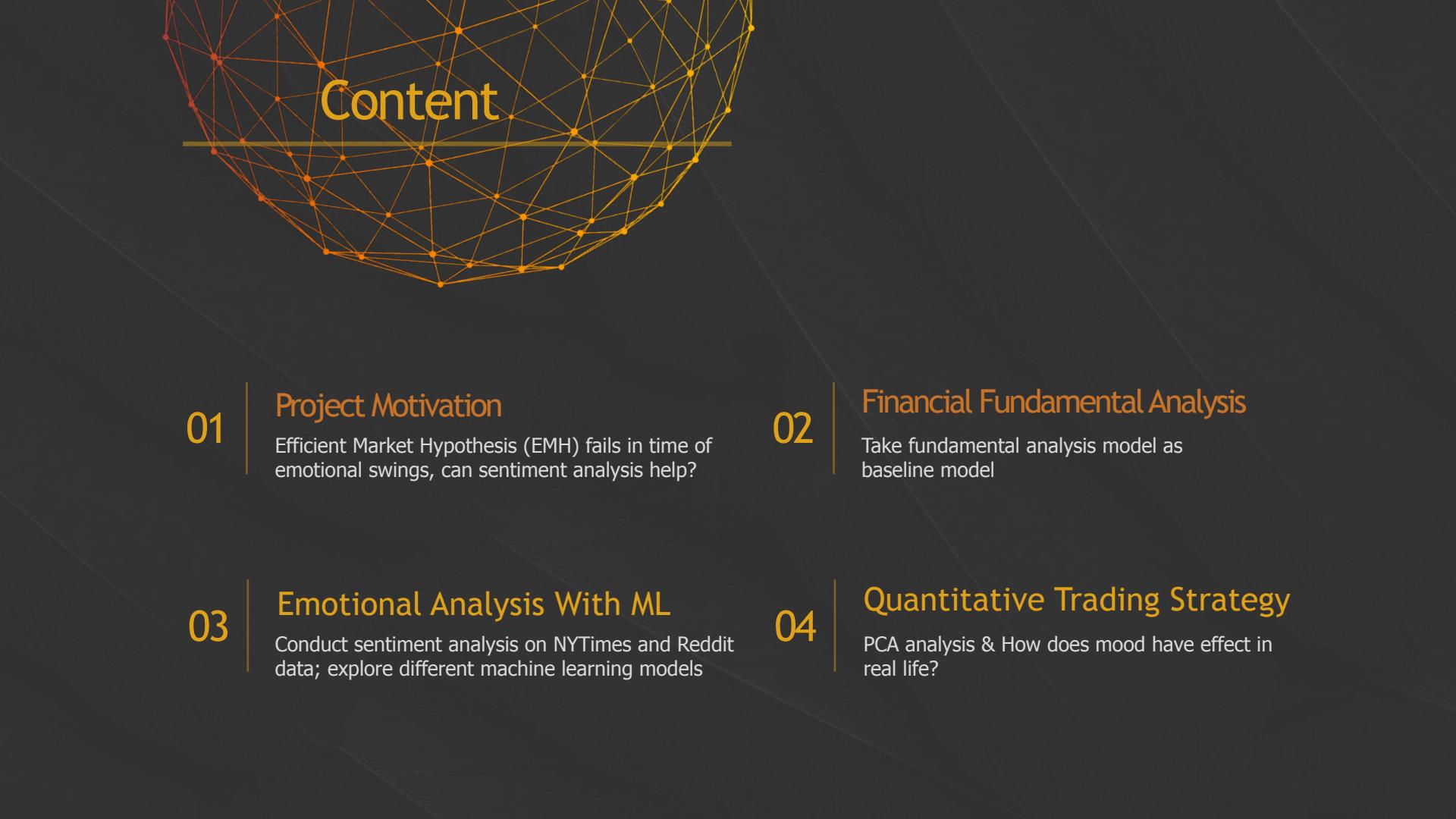




Stock Prediction & Trading using Social Media Sentiment

AnalyticsQueenBee Group



Content

01

Project Motivation

Efficient Market Hypothesis (EMH) fails in time of emotional swings, can sentiment analysis help?

02

Financial Fundamental Analysis

Take fundamental analysis model as baseline model

03

Emotional Analysis With ML

Conduct sentiment analysis on NYTimes and Reddit data; explore different machine learning models

04

Quantitative Trading Strategy

PCA analysis & How does mood have effect in real life?



01

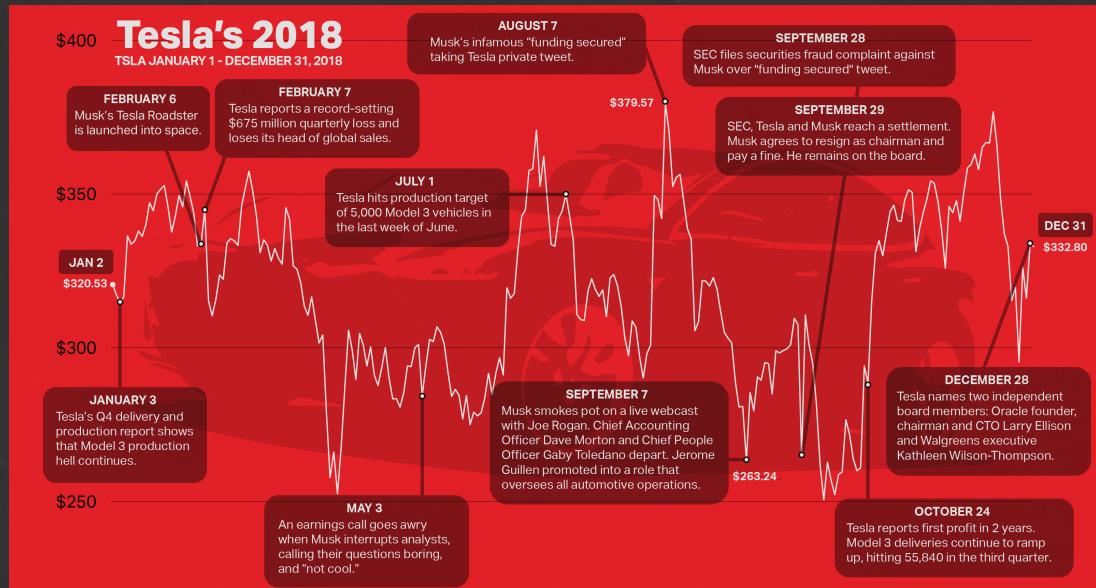
Project Motivation

Efficient Market Hypothesis (EMH) FAILS in time of emotional swings, can sentiment analysis help?

Project Motivation - EMH Failure

Tesla Case - the privatization drama

2018.8.7



Project Motivation - EMH Failure

Facebook Case - Data Scandal



Rough year for Facebook shares

FB stock price, January 2018-present



Project Motivation - EMH Failure

Facts:

Institutional stock valuation have failed in predicting prices and helped little in trading these major stocks



Turn to Behavioral Finance:

It assumes market is INEFFICIENT and Mood is a kind of behavioral biases.

Question:

What are possible emotion proxies?

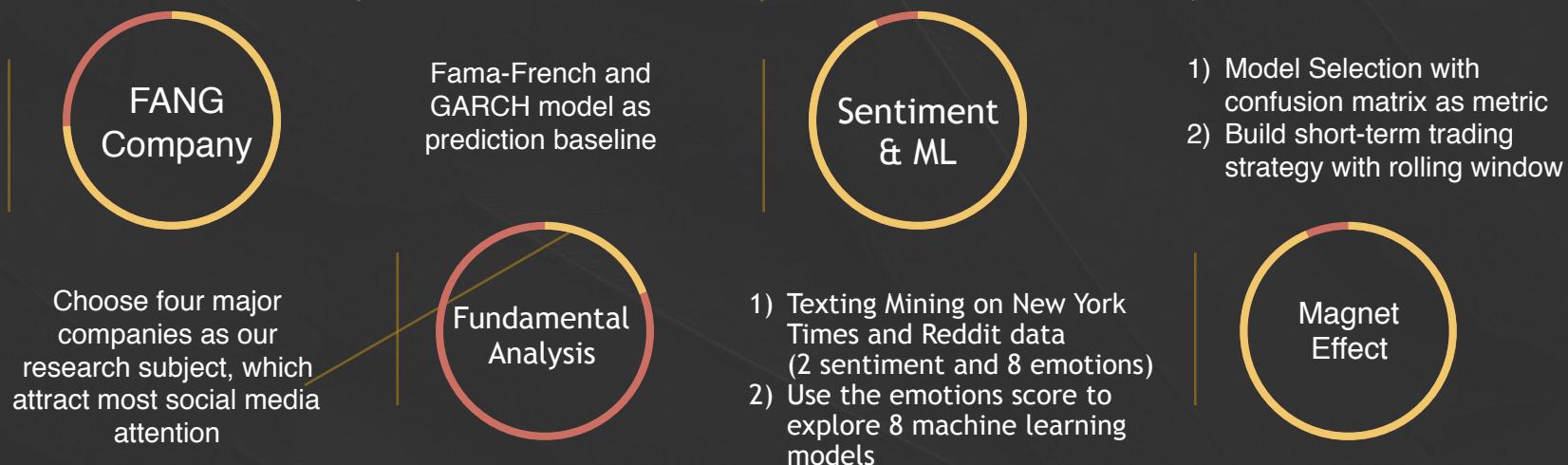
How could sentiment analysis serve good for investors on these major stocks?

Project Motivation - EMH Failure

Data

{Company: Facebook, Amazon, Netflix, Google,
Stock Price: 1 year stock data,
Sentiment Proxy: New York Time Article (institutional voice)
Reddit Discussion under Economy section (public opinion)

Model





02

Fundamental Analysis

How do traditional methods evaluate the stock?

- Fama French
- GARCH

2.1 Fama French Three Factor Model

The Fama French 3-factor model is an asset pricing model that expands on the capital asset pricing model by adding size risk and value risk factors to the market risk factors.



SMB
Small Minus Big
Small-cap stocks tend to outperform large-cap stocks.



$R_{Mt} - R_{ft}$
Market Excess Returns



HML_t
High Minus Low
Value stocks tend to outperform growth stocks.

2.1 Fama French Three Factor Model

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \epsilon_{it}$$

R_{it} is the total return of a stock or portfolio, i at time t

R_{ft} is the risk free rate of return at time t

R_{Mt} is the total market portfolio return at time t

$R_{it} - R_{ft}$ is expected excess return

$R_{Mt} - R_{ft}$ is the excess return on the market portfolio (index)

SMB_t is the size premium(market capitalization)

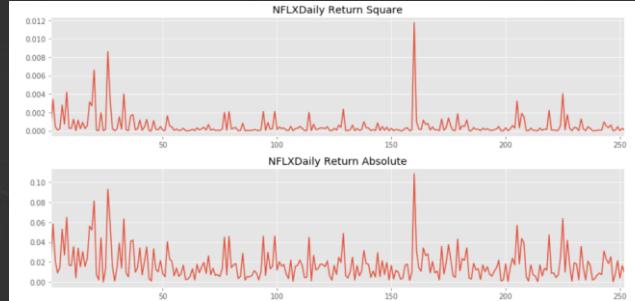
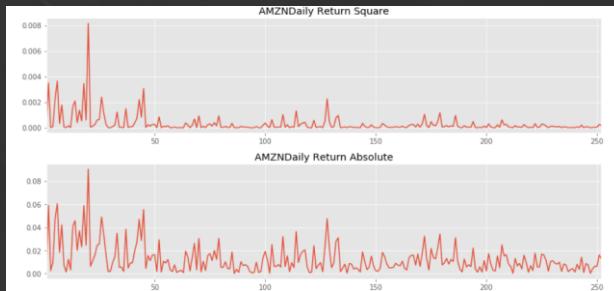
HML_t is the value premium(book value to market value ratio)

$\beta_{1,2,3}$ refer to the factor coefficients

	Accuracy
GOOG	0.74
FB	0.59
AMZN	0.73
NF	0.94

2.2 GARCH Model

The graphs show that return series have some properties such as fat tail and volatility clustering.



2.2 GARCH Model



01

Volatility Attributes

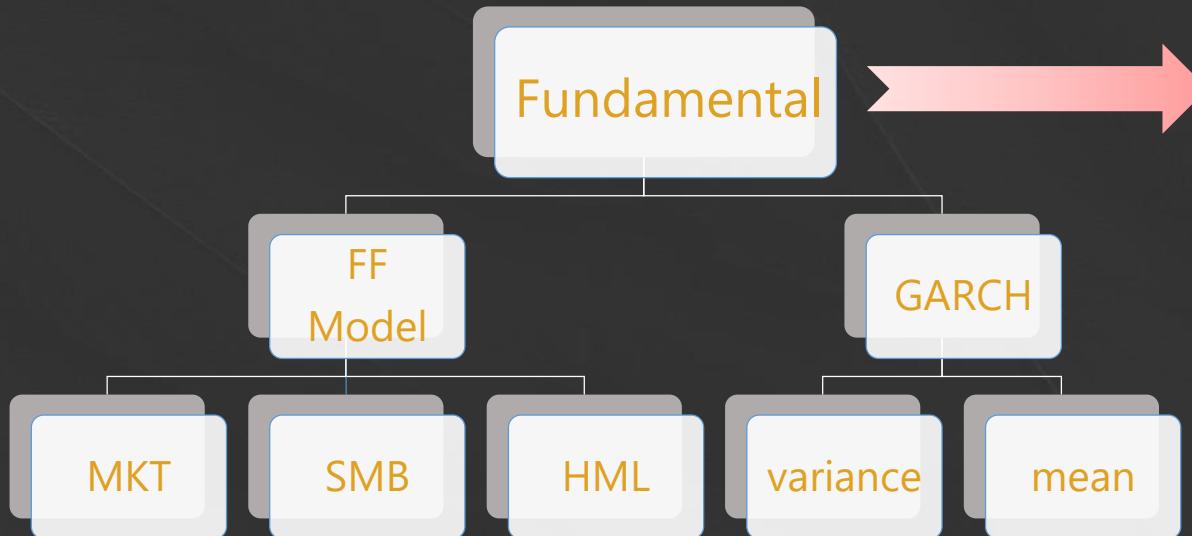
Volatility becomes more volatile during financial
and less volatile during steady economic growth.

02

Volatility Prediction

Adding Apply the GARCH model to predict the stock
price. The accuracy is around 40%.

2.3 Conclusion



Well-exploited financial models
Do not necessarily yield good
Results.

Once the market sees the
pattern, the pattern will
disappear.
Baseline to test other
approaches.



Emotional Analysis With ML

Conduct sentiment analysis on NYTimes and Reddit data; explore different machine learning models

Emotional Analysis With ML - Getting Data

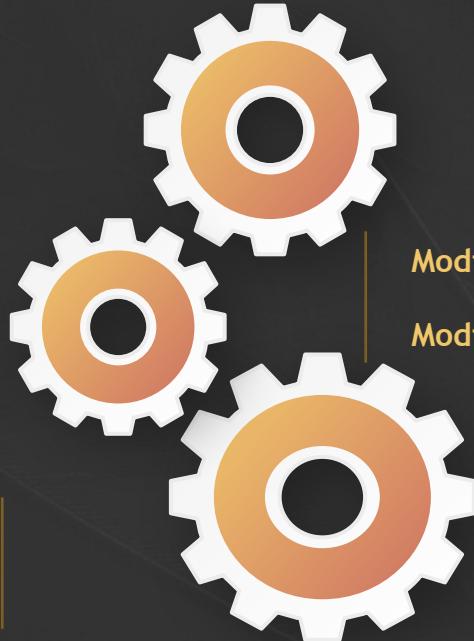
For Every Company:

New York Times Article API

1. Article Title
2. Article Abstract
3. Date

Reddit API

1. Upvote
2. Score
3. Discussion number
4. Discussion content
5. Date

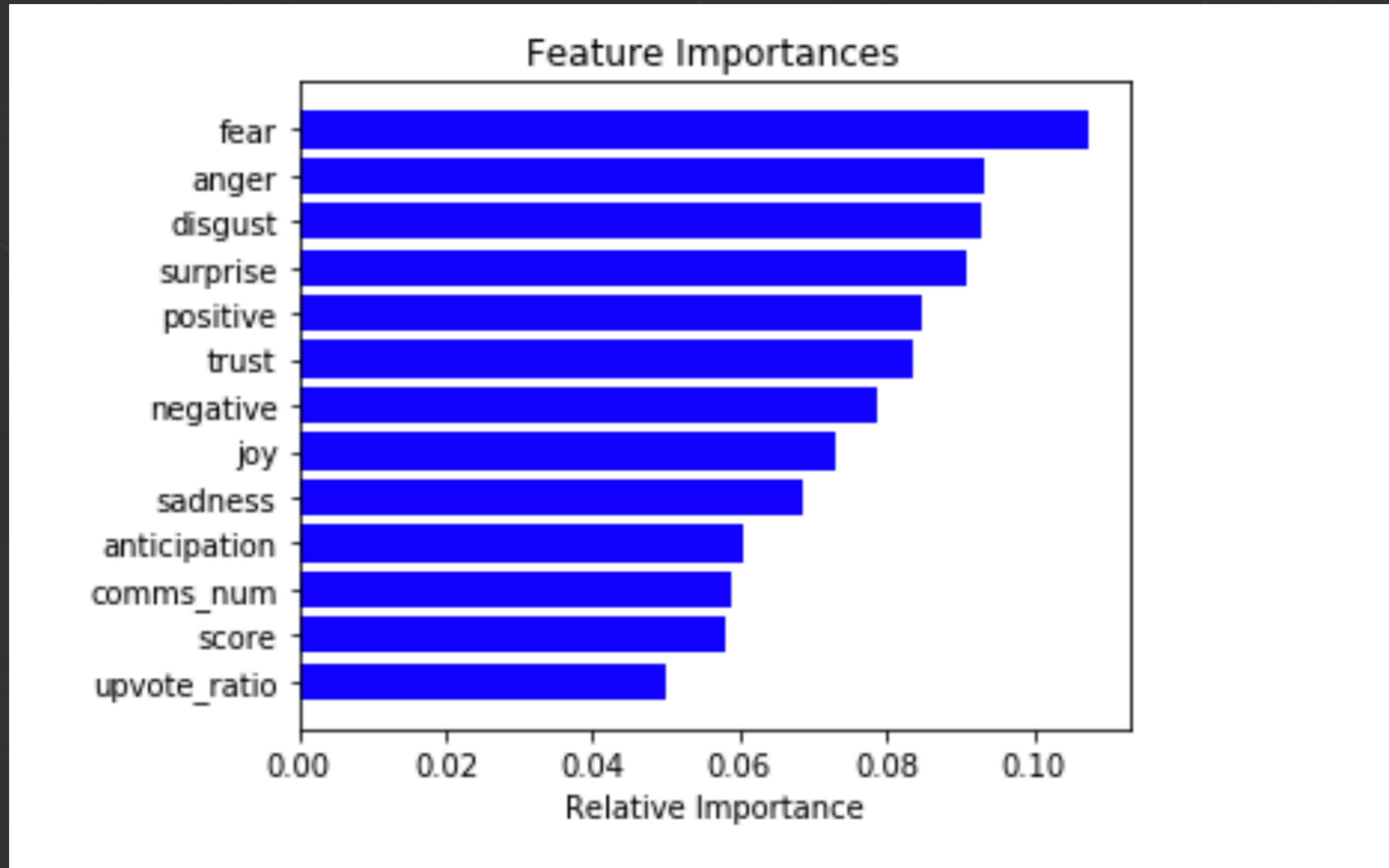


Text Mining with NRC Lexicons

Modified Positive = Positive + joy + trust + anticipation

Modified Negative = negative + disgust + anger + fear

Emotional Analysis With ML - Emotion Importance



Emotional Analysis With ML - Model Exploration

- 1. linear regression model
- 2. Decision tree: Classification trees
- 3. Random forest
- 4. SVM
 - (1) linear SVM
 - (2) non-linear SVM: sigmoid
- 5. Neural Network
 - (1) linear NN: 1 layer
 - (2) non-linear NN: 2-3 layers
- 6. Logistic Regression
- 7. Xgboost(too complicated, overfitting)

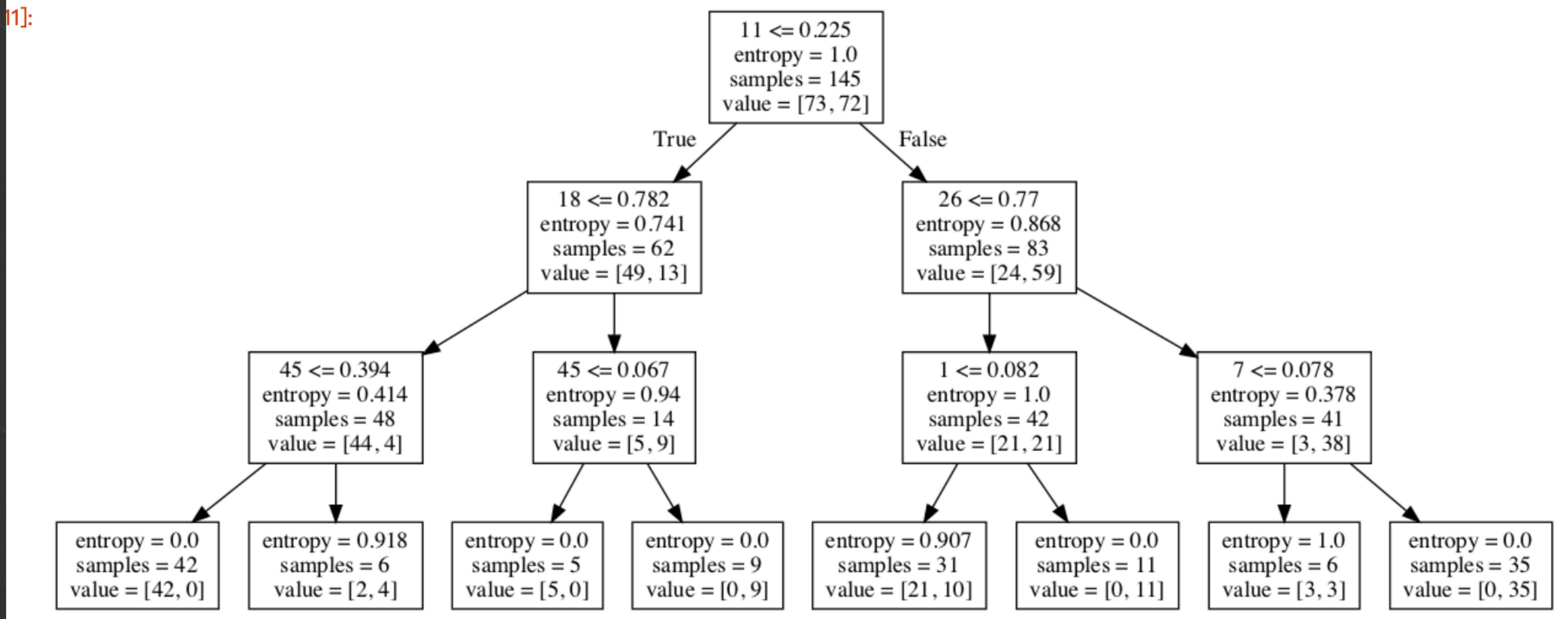
Binary Classification of 0 & 1

Randomly split train and test: 7:3

Reddit data has strong precision

Emotional Analysis With ML - Model - Decision Tree

11]:



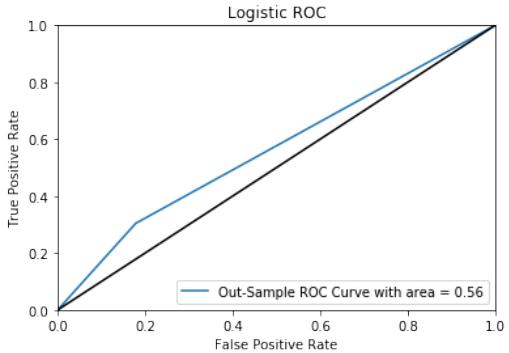
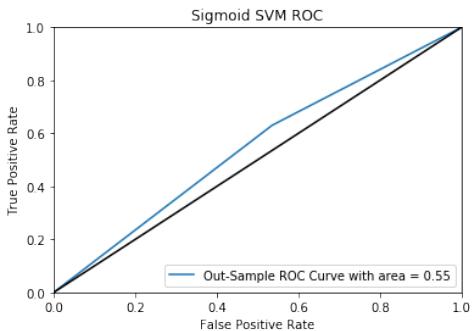
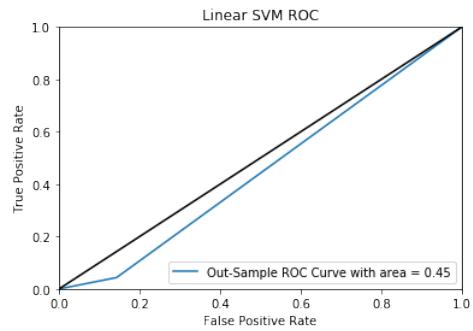
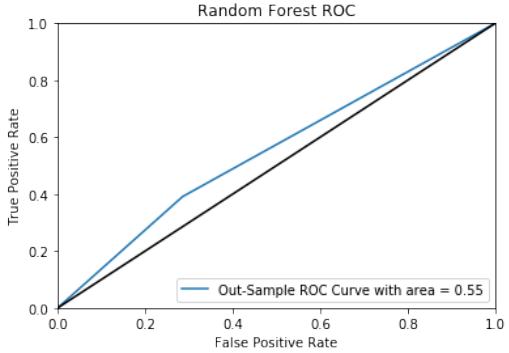
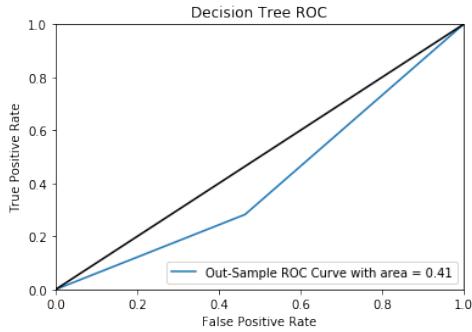
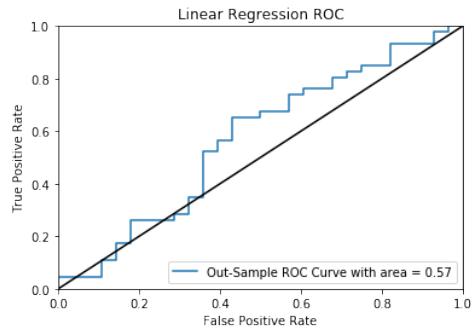
Emotional Analysis With ML - Model Exploration

Binary Classification

Reddit data has strong precision
Randomly split train and test: 7:3

	Precision	True_Negative_Rate	False_Positive_Rate	f_score	Accuracy
Lnear Regerssion	0.571429	0.410256	0.342857	0.477612	0.527027
Decision Tree	0.500000	0.666667	0.742857	0.571429	0.472973
Random Forest	0.533333	0.410256	0.400000	0.463768	0.500000
Linear SVM	0.527027	1.000000	1.000000	0.690265	0.527027
Non-linear SVM(sigmoid)	0.641026	0.641026	0.400000	0.641026	0.621622
Linear Neural Networks	0.527027	1.000000	1.000000	0.690265	0.527027
Non-Linear Neural Networks	1.000000	0.051282	0.000000	0.097561	0.500000
Logistic~	0.530303	0.897436	0.885714	0.666667	0.527027

Emotional Analysis With ML - Model Exploration



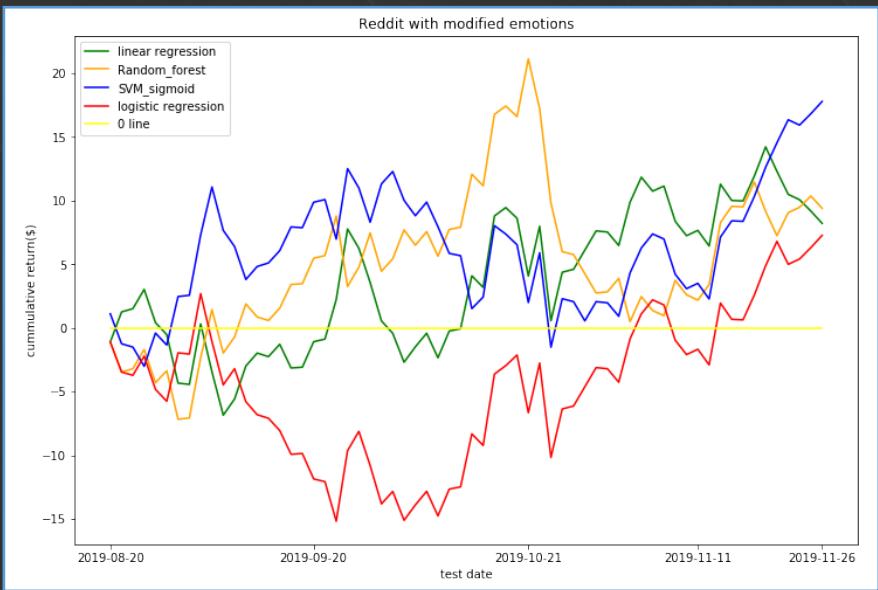
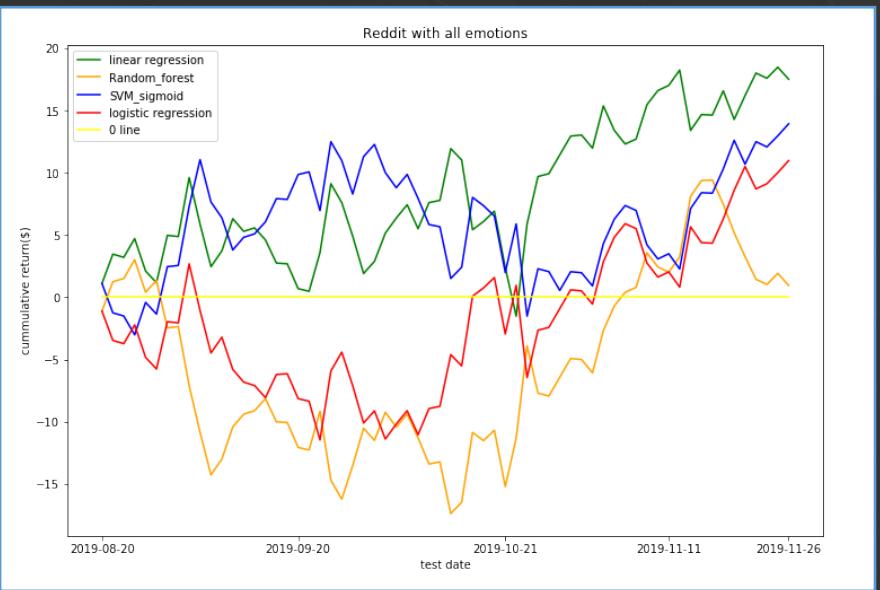


04

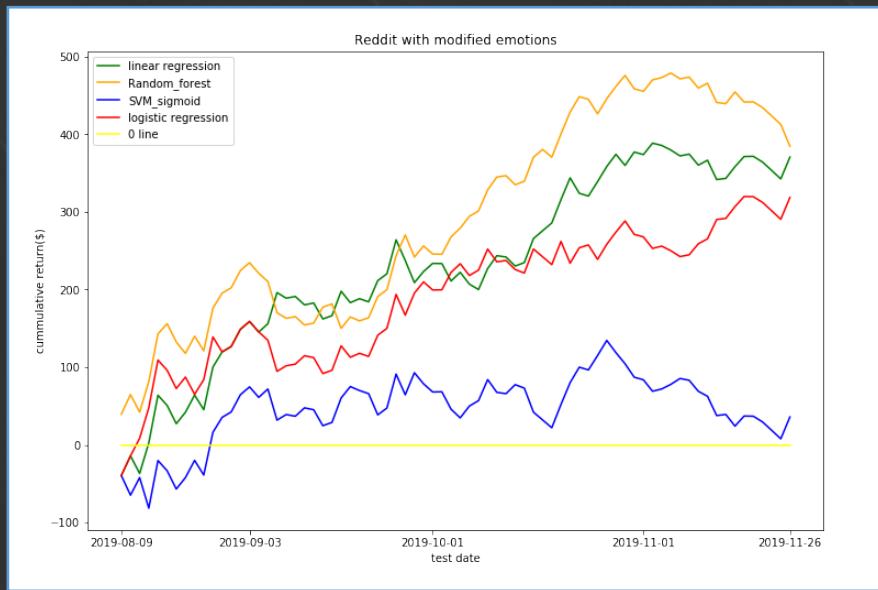
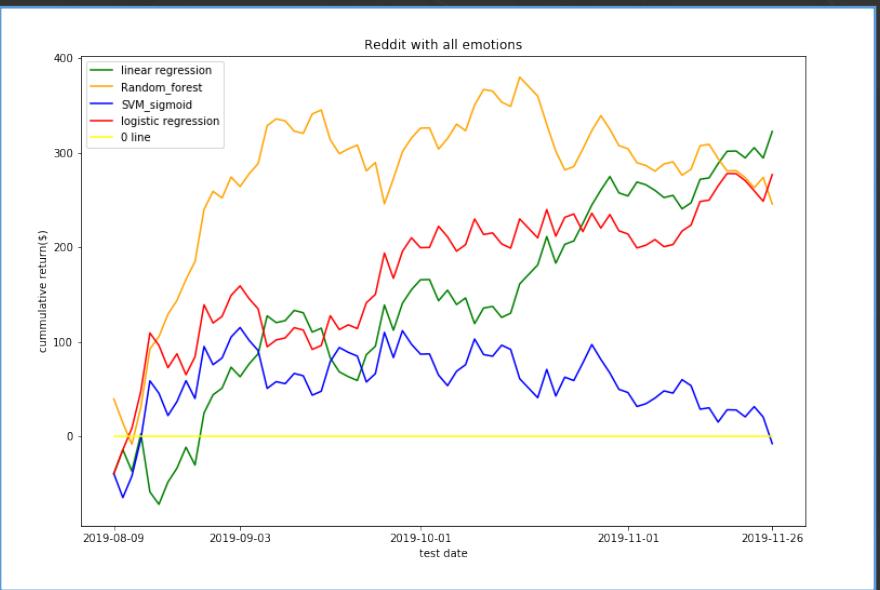
Short-Term Trading Strategy

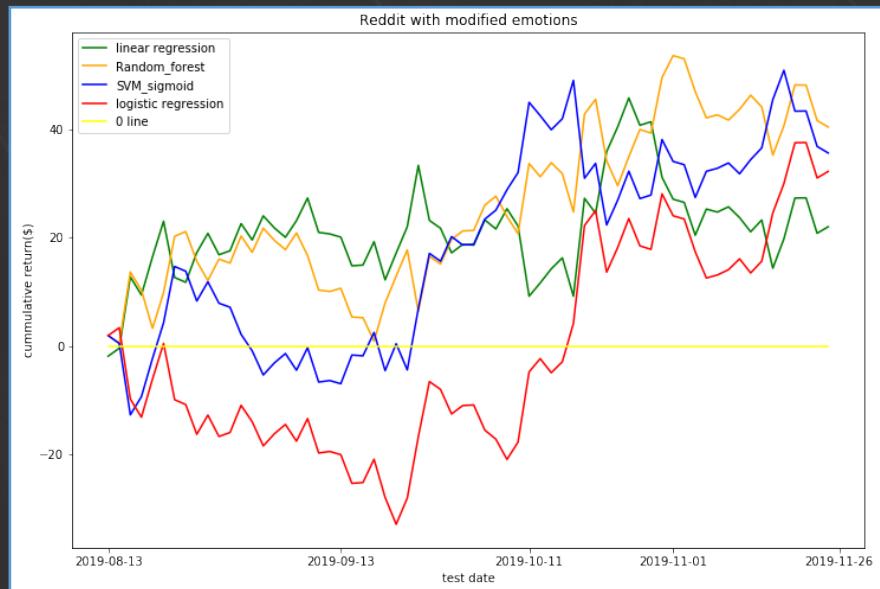
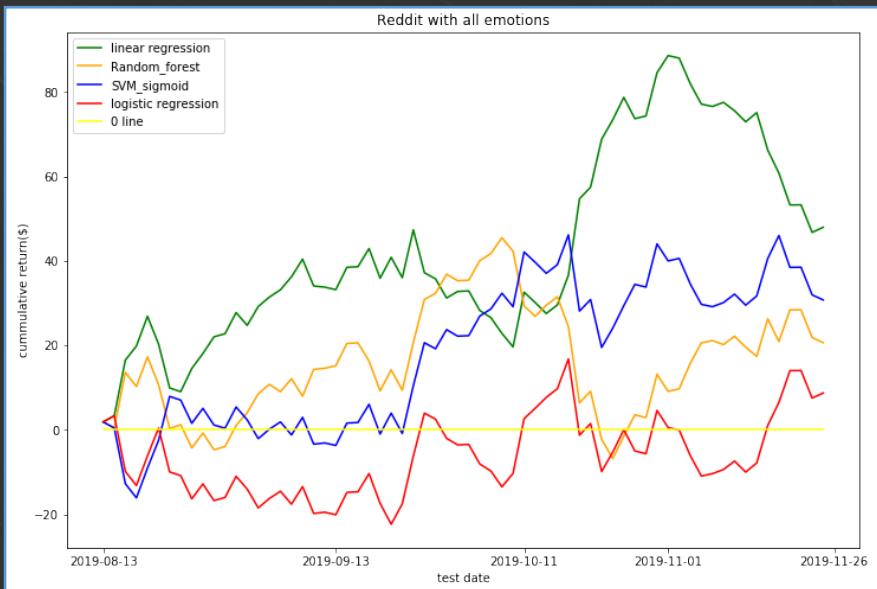
How to used these ML model to build quantitative trading strategy, what models performs best

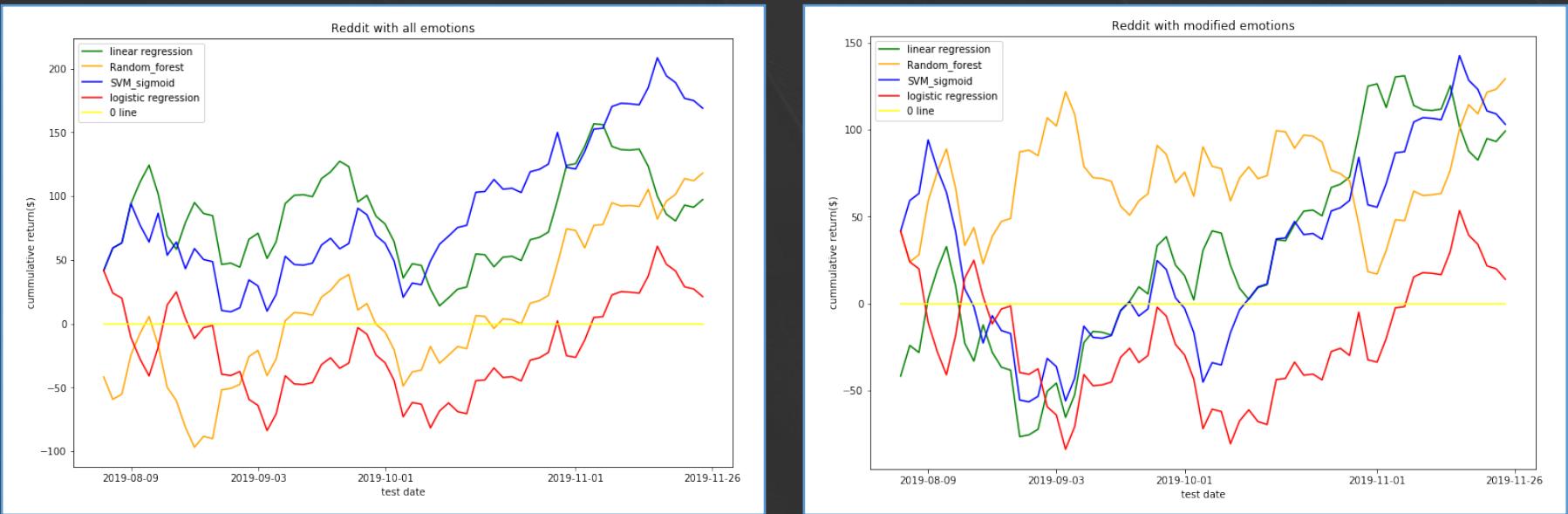
Facebook



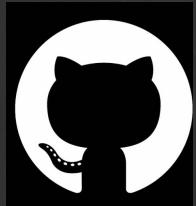
Amazon







Our Repo



Welcome to visit our github repo:

https://github.com/danielle707/stock_sentiment_analysis_using_social_media.git



MANY THANKS