Deutsche Bank
Corporate Banking & Securities

News Analytics

Presented by Jayanth, Summer Intern, Rates & Equities

Passion to Perform

Project sponsor:
Romith Das, Rates

Project guide:
Adithya MR, Rates Trading

Mentors:
Dinesh B Goel, Equities
Ankit Tiwari, Rates Trading

Contents

Introduction

About Project

Background

- News convey information, which lead different investors to make decisions
- Above decisions impact the market, which aggregates the information that investors receive and reflects it via the price
- Advances in Natural Language Processing has made it possible to quantify the news “sentiment” in real time, opening up a whole new data source for analysts

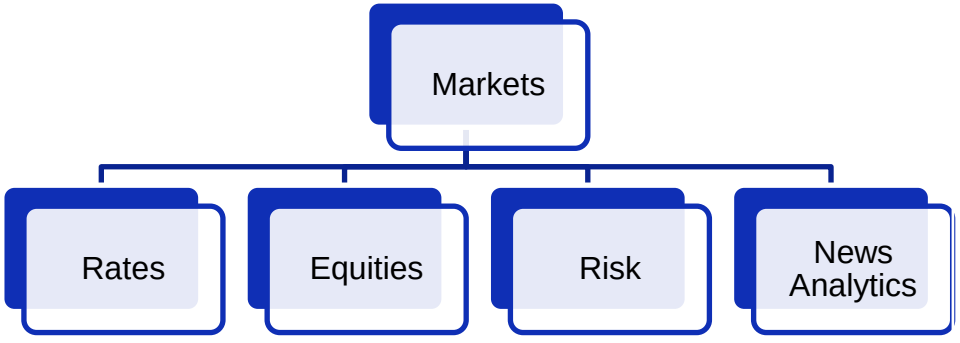
Why did I choose the project?

- Multi domain project with utility in different verticals
- Application of Machine Learning in Finance (Applied Computer Science)

Objective

- Web scrape asset specific news data
- Summarize news and extract keywords
- Provide ‘Sentiment Score’ to the news
- Predict stock price movement based on news & build an index

Beneficiaries

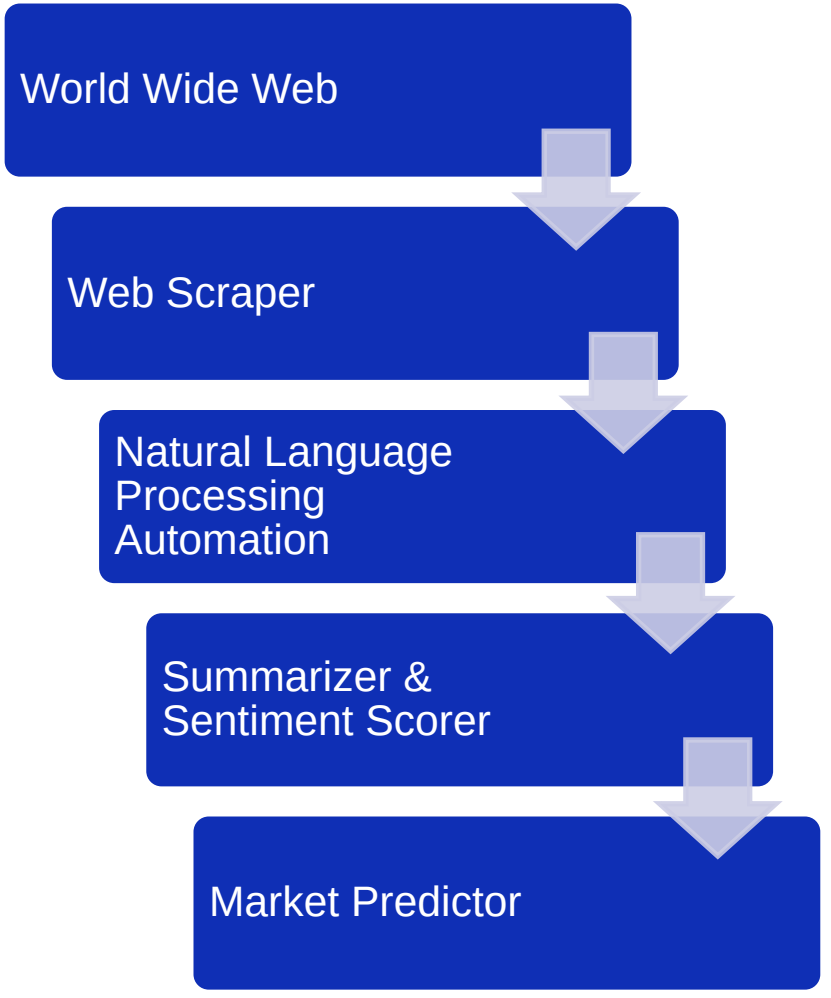


Project Outline

Approach

- Level 1: Building Web Scrapper
  - ✓ Made Feed Downloader and Parser to get news feed with content
  - ✓ Made Ticker based News Downloader
- Level 2: Building & Testing Natural Language Processing Tools
  - ✓ Entity Extractor
  - ✓ Rapid Automatic Keyword Extractor
  - ✓ Summarizer
  - ✓ Dictionary Based Sentiment Scorer
  - ✓ Supervised Sentiment Scorer
- Level 3: Market Prediction & Index Development
  - ✓ Direction Prediction Model
  - ✓ Trading Strategies

Overall Picture



Cleaning & Filtering

Keyword Extractor : It extracts the key phrases from the news by first identifying the keywords based on their frequencies and then, merging them.

Entity Extractor : It selects the “Proper Noun” from the news content after Parts of Speech tagging so as to identify the stakeholders in the news.

Domain Classifier : It classifies the news as “Business” news or “General” news, for filtering out the noise from scraped news data.

Summarizer : It extracts key sentences based on keyword frequencies and the recommendation of sentences by other sentences in the news .

Title
- Economics Not the Only Issue With PetroChina IPO

Domain
- Business

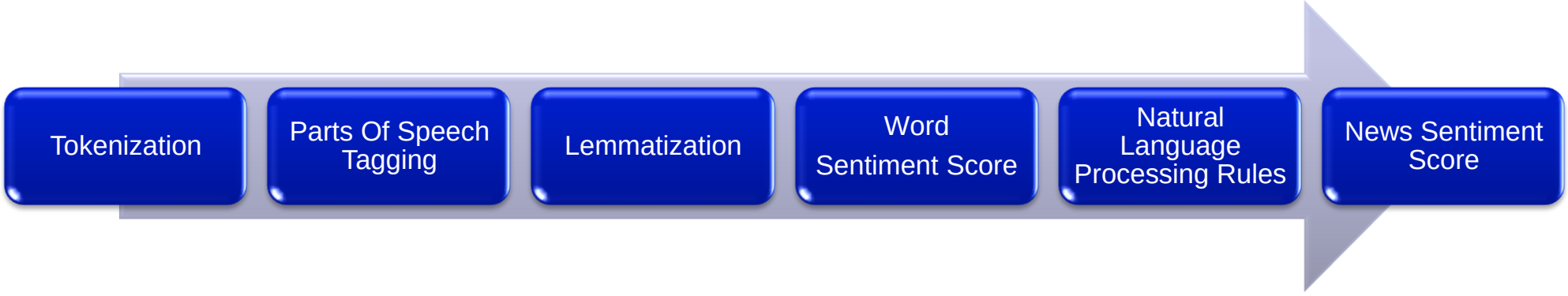
Entities
- China, BP Amoco, PetroChina

Keywords
- State-owned parent company would meet BP Amoco 's standards
- Two companies would jointly market fuels

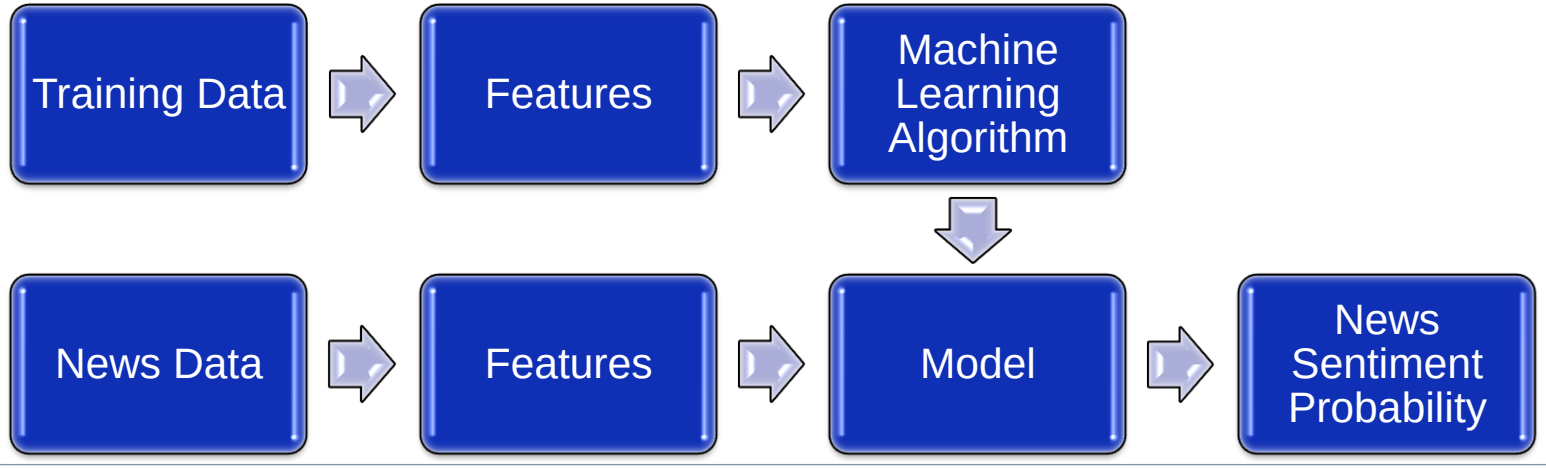
Summary
- Stepping into controversy, BP Amoco said Thursday that it would take a 20% stake in the initial public offering of PetroChina.

Sentiment Analysis

1. **Natural Language Processing Based Model**: The model assigns a score to each word based on Loughran and McDonald Financial Word List, consisting of
    - Positive and negative,
    - Sentiment incrementing and decrementing,
    - negation word list.

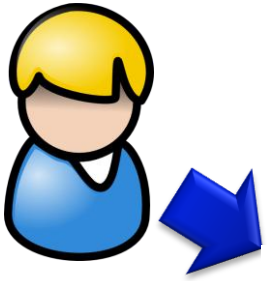


2. **Machine Learning Model**: The model uses Naïve Bayesian Classifier, trained on the manually annotated news data from GTB Risk Group.



Trading Strategies

Natural Language Processing Analyst



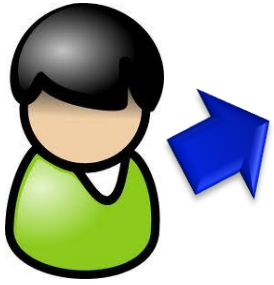
VOTE

Possible sentiments

Positive : 1

Negative : -1

Neutral: 0



Machine Learning Analyst

SIMPLE MOVING AVERAGE

Average the sentiment values over the last N days

DISSOLVING SENTIMENT

Wait for sentiment to dissipate to zero over recent history

EXPONENTIAL MOVING AVERAGE

Take exponential mean of the previous sentiments to account for eventful or tail risk days

Positive: Long
Negative: Short
Neutral: Hold

Leverage can be increased using the number of votes to a particular signal to get more returns, e.g. investing in more asset quantity based on stronger sentiments

Strategy Comparison

Google (GOOG US EQUITY) : August 2004 – June 2013

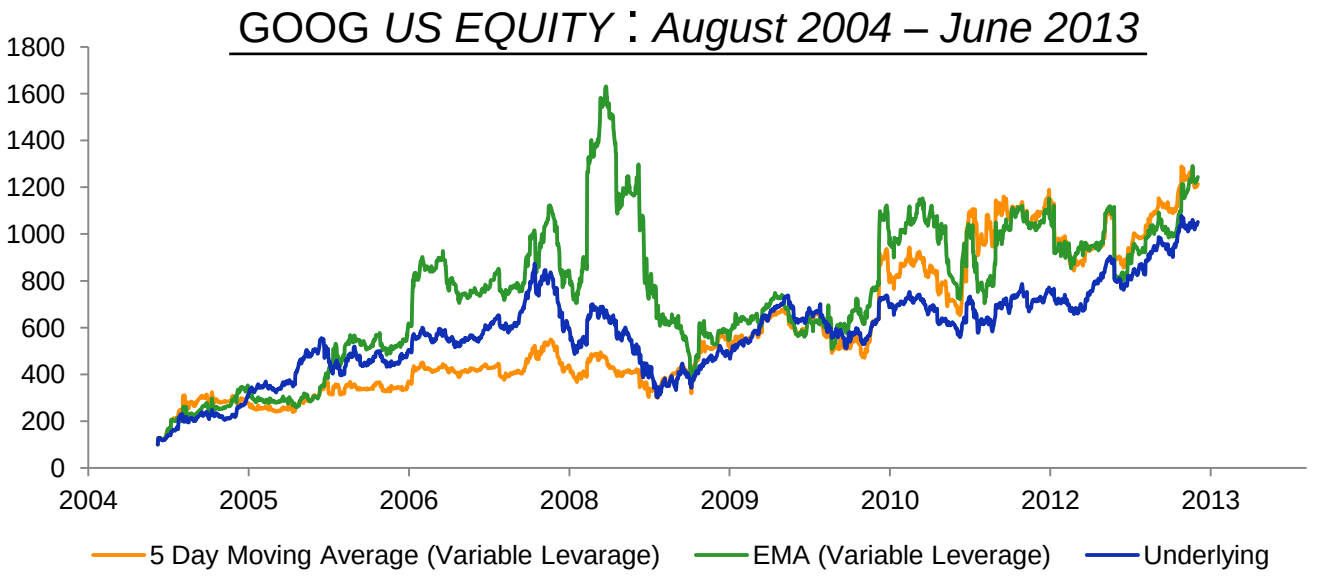
	Annualized Return	Volatility	Sharpe Ratio	Maximum Drawdown
Underlying	30.34%	34.45%	0.881	-65.29%
Daily Sentiment	17.94%	31.18%	0.575	-66.73%
Daily Sentiment (Variable leverage)	25.87%	57.26%	0.452	-89.34%
5-Day Sentiment Moving Average	29.26%	33.51%	0.873	-52.72%
5-Day Sentiment Moving Average (Variable leverage)	32.46%	37.68%	0.862	-45.36%
Exponential Moving Average (Variable leverage)	32.83%	47.94%	0.685	-77.82%
Dissolving Sentiment	21.97%	31.77%	0.692	-54.02%

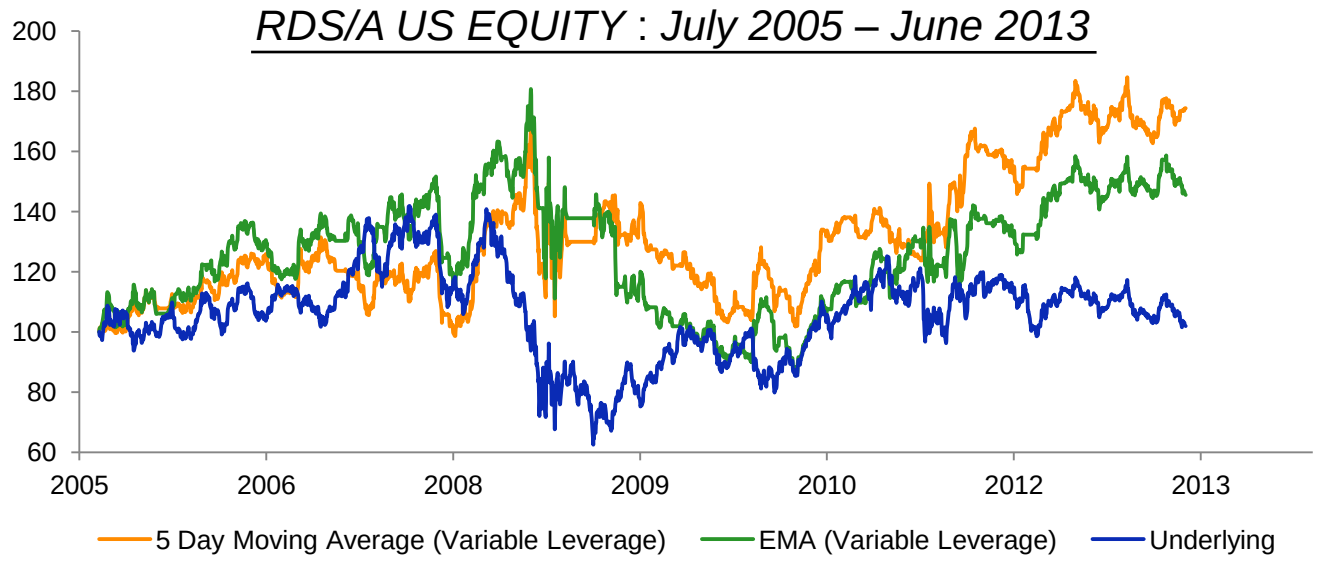
Royal Dutch Shell PLC (RDS/A US EQUITY) : July 2005 – June 2013

	Annualized Return	Volatility	Sharpe Ratio	Maximum Drawdown
Underlying	0.24%	29.76%	0.008	-55.90%
Daily Sentiment	5.58%	25.40%	0.220	-45.50%
Daily Sentiment (Variable leverage)	5.11%	27.31%	0.187	-52.25%
5-Day Sentiment Moving Average	3.41%	27.27%	0.125	-40.58%
5-Day Sentiment Moving Average (Variable leverage)	7.23%	24.34%	0.297	-39.21%
Exponential Moving Average (Variable leverage)	4.83%	26.10%	0.185	-52.30%
Dissolving Sentiment	4.10%	26.08%	0.157	-52.72%

5-Day Sentiment Moving Average with variable leverage seems to be the more consistent strategy. Both Exponential Moving Average and 5-Day Sentiment Moving Average outperforms the underlying.

Index Performance





Conclusion

1. **Results**
    1. The strategy works best on high frequency and highly volatile assets as they readily absorb the news sentiment and reflect it via their price.
    2. The effect of the news sentiment decreases exponentially with time and it vanishes as soon as within five business days for highly volatile equities.
    3. Exponential Moving Average and Simple 5 Day Moving Sentiment Average are the better market predictors.
    4. The tools built within this project framework allows us to process an unlimited number of news stories, essentially in real-time, easing the job of reading and interpreting new information as it is released in the market.

2. **Areas of improvement**
    1. The sensitivity to search methods is high, and web crawling is not an optimal route for this.
    2. The news signal are only able to predict the trend but not the mean reversion and thus, the strategy must be employed with mean reverting strategies.
    3. Most of the news articles are not future market signals but rather, mirror of past market movements. To take care of the false signals, Moving Average Price Convergence and Divergence must be used to recheck the news signals.
    4. The sentiment probability can be taken into account for leverage based strategies.
    5. All the results so far have been pre-transaction costs. Transaction Cost must be taken into account to get net returns.