BlackRock

Stock Directionality Using Alternative Data

Themes & Data

Theme

Social Sentiment Based Models

Data

- arXiv Publications (Scraped)

Quandl

- Social Media Analytics (SMA1)
- End of Day US Stock Prices (EOD)
- Wiki Continuous Futures (CHRIS/ICE_B28)

‘Let’s identify the strongest directional signals for public equities’

BlackRock®

Our Hypothesis

Academic Publication Volume As A Feature

Quantifying the macro economic trend for machine learning services judged by the hype of the public and academic research

Social Media Data and Sentiment Analysis

Long term S&P share directionality prediction from social media data for individual companies

S&P Oil Companies Sentiment informing the price of Brent

Reinforcement Learning Market SImulations

Taking a many agent, central exchange, reinforcement learning approach to better backtesting our trading strategies.

BLACKROCK

1. S&P Oil Company Social Media Signals To Predict Brent

Model Features

Aggregating 29 S&P listed oil firms

Instagram, Twitter and Facebook signals

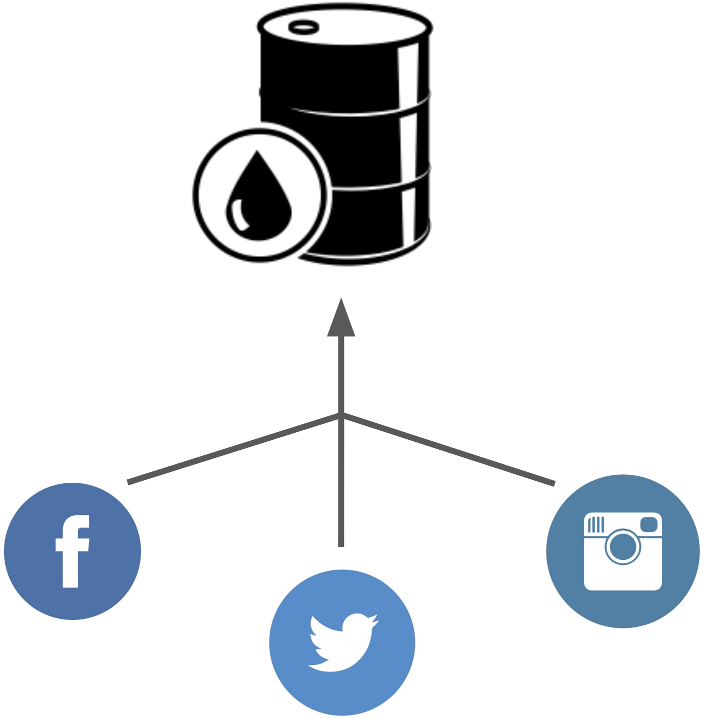
Random Forests classifier trained on daily data - 55% prediction accuracy

Model Learnings

Next day predictions are extremely noisy

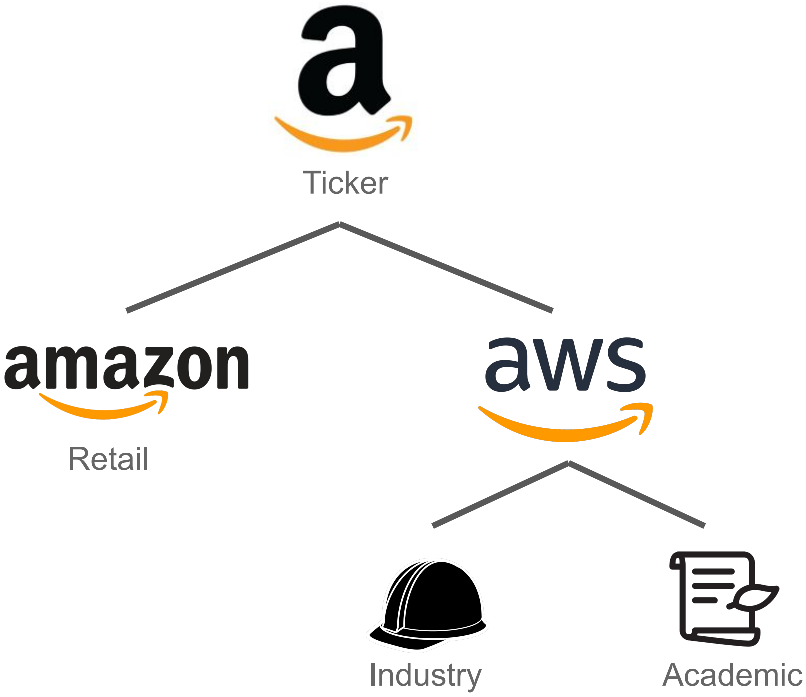
Single company predictions are better

Social media signals are significant



BlackRock

2. Machine Learning Momentum - AWS Usage Influencing AMZN Price



Academic Research

Breadth of academic research indicates the extent of future machine learning usage in industry

AWS Revenue Feeds AMZN

An uptick in AWS usage heavily influences the AMZN stock price

3. Modelling Trends WIth Social Media Data

Longer Term Directionality Prediction

Predict rise or fall in share price from the current point at a 30 day interval

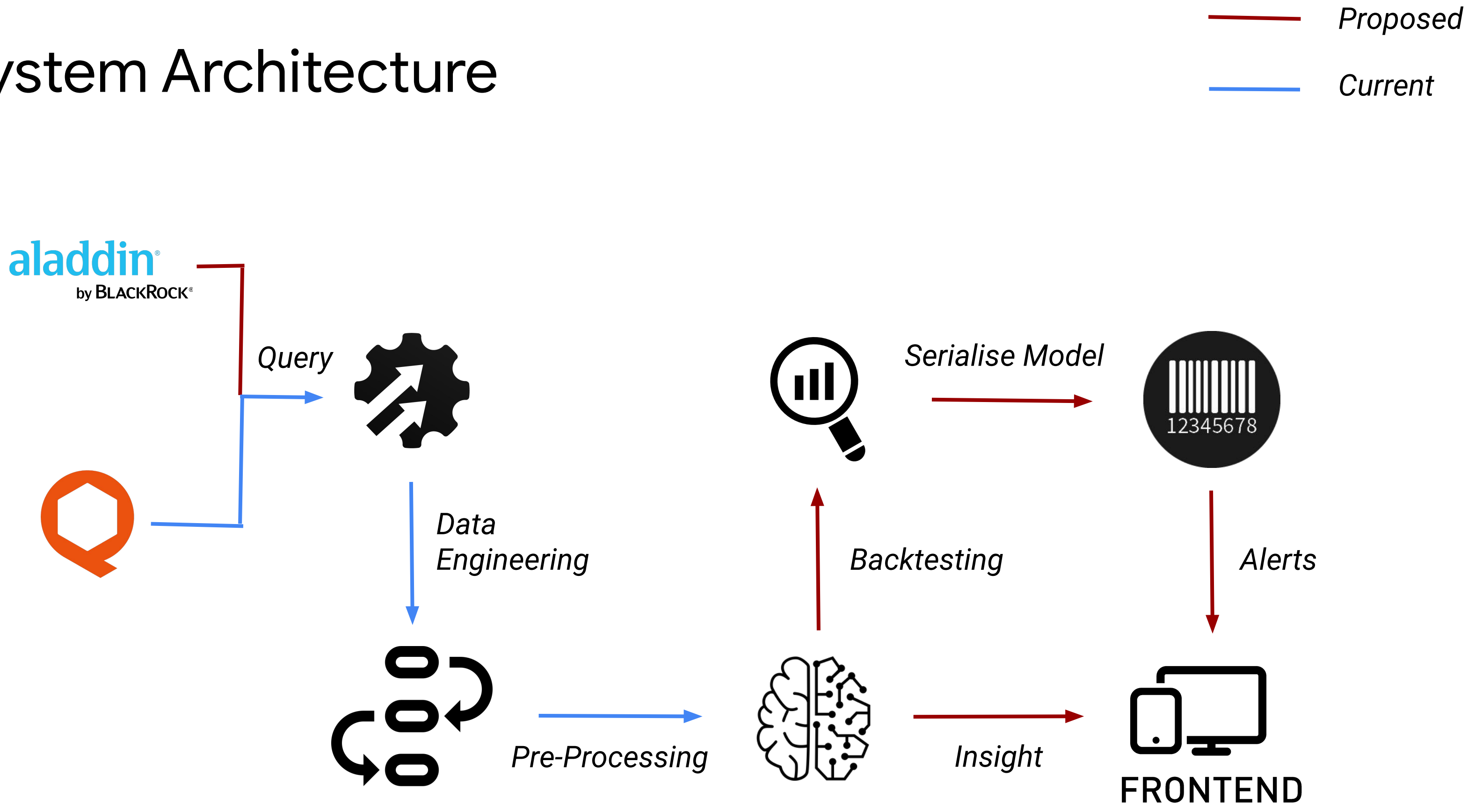
Individual Firm Level Prediction

Time series prediction at the specific firm level

Social Media Signals

Utilising Facebook and Instagram posts - shares, comments, likes, reactions and sentiment

System Architecture



Model Performance - Classification Accuracy

90.3%	**91.6%**	**86.7%**	**87.9%**
MCD	AMZN	NFLX	NKE
$\begin{bmatrix} 40 & 10 \\ 6 & 109 \end{bmatrix}$	$\begin{bmatrix} 41 & 8 \\ 6 & 12 \end{bmatrix}$	$\begin{bmatrix} 60 & 15 \\ 7 & 82 \end{bmatrix}$	$\begin{bmatrix} 52 & 8 \\ 12 & 93 \end{bmatrix}$

BLACKROCK

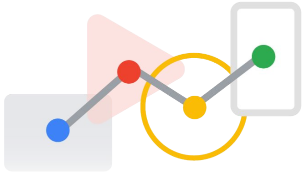
Isn’t this a bit high?

Quandl Data Leakage

Actions and Sentiment from during and after the date of posting are included in data

Leakage Effect

Additional future information can better correlate with stock price changes, artificially increasing model performance



Limited Exposure

We predict 30 days forward using the previous months data, almost all actions and collected sentiment should have been captured previous to that point

Why Might This Model Work?

Highly Relevant Trading Signals

Volume of activity and sentiment massively affect consumer facing businesses

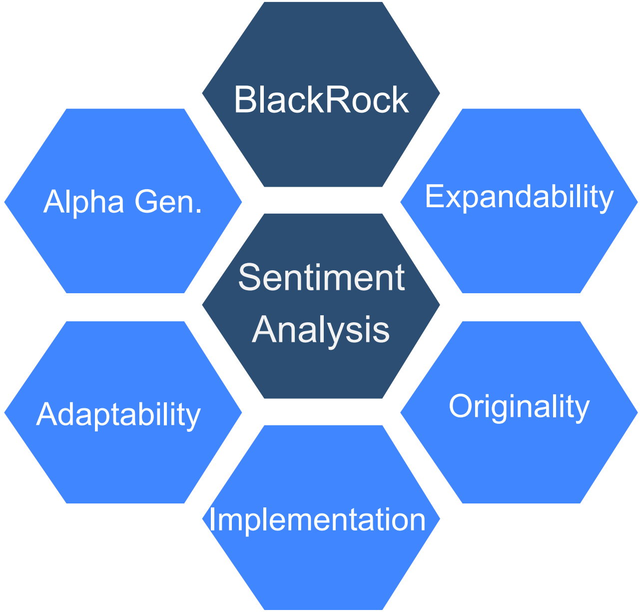
Long Range Predictions

Forward monthly predictions overcome random short term noise

Single Company Focus

One Random Forests model trained per firm

Model Advantages



Alpha Generation

Positions on a month long trading horizon

Adaptable Performance

Long term adjustability depending on preceding conditions

Expandability

Applicable to all consumer focussed firms with sentiment and social media data available

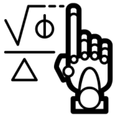
Originality

Using niche features such as facebook reactions in addition to sentiment analysis

Implementation Viability

Scalable across Aladdin datasets

4. Realistically Simulating Our Strategy - RL



Reinforcement Learning Simulations

An Improvement to Monte Carlo methods for backtesting trading strategies

Thousands of agents with their own reward functions interacting with a central exchange

Limitations

Requiring level 3 data market data to train many agents to interact with central exchange