Bitcoin Trade Recommender

Phat Doan, Nate Velarde, Linh Tran, Raymond Lee

Project Objectives

Use Cases

- **Research Question:** Can social media improve the performance of Price Momentum Strategies?
- **Algorithm Analysis & Optimization:** using historical and live data, adjust signals and trading window to optimize trading algorithm

Data Architecture

- **Storage & Streaming:** merge various sources (CSV, API) for processing
- **Processing:** parse and aggregate tweet data, merge with BTC data, run hypothetical trading scenarios
- **Serving:** filter most pertinent records & attributes for trading performance analysis

Can a Social Media Factor Improve the Performance of Price Momentum Strategies?

- **Base Price Momentum Trading Signal**
  - BUY when Current Price > 5 Day Moving Average
  - SELL when Current Price < 5 Day Moving Average

- **Social Media Factor - Bitcoin Tweets**
  - Basic Signal
    - Bitcoin Tweet Counts
  - Advanced Signal
    - Ratio of Positive to Negative Tweets as determined by sentiment analysis

Can a Social Media Factor Improve the Performance of Price Momentum Strategies?

- **Basic Signal**
  - BUY when Current Tweet Count > 5 Day Moving Average
  - SELL when Current Tweet Count < 5 Day Moving Average
- **“Advanced” Signal**
  - BUY when Ratio of Positive to Negative Tweets > 5 Day Moving Average
  - SELL when Ratio of Positive to Negative Tweets < 5 Day Moving Average

Can a Social Media Factor Improve the Performance of Price Momentum Strategies?

- **Combination Strategy - Base Price Momentum Signal + Tweet Signal**
  - 100% invested when both Price and Social Media signals are BUY
  - 50% Invested when Price and Social Media signals conflict
  - 0% invested when both Price and Social Media signals are SELL

Social Media as a Signal

Tweets Count

Current strategy

- Number of tweets related to bitcoin per day (~350k/day)
- Ratio of bitcoin tweets over all tweets (~1%)

Text Analysis

Enhanced strategy

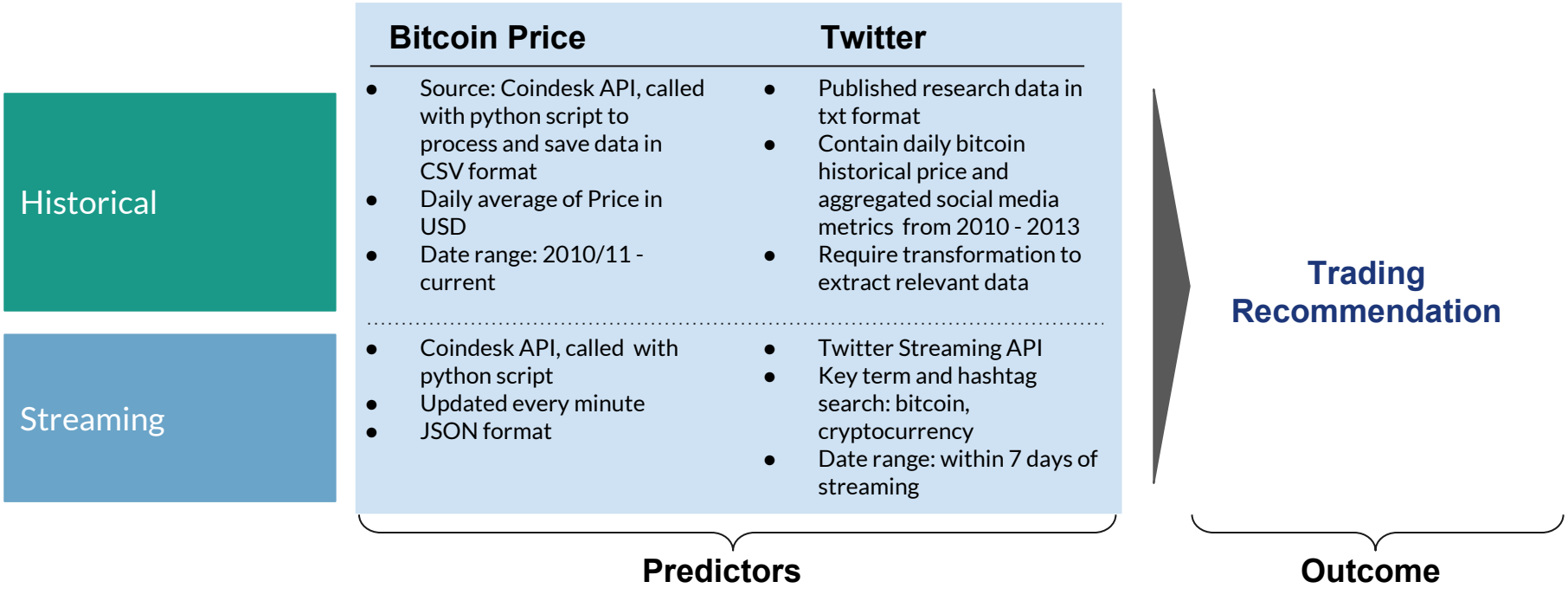
- Remove stopword, hyperlink, and non-English words
- Combine all tweets within a specific period together
- Vectorize and conduct Logistic Regression to find non-zero feature (important) words
- Ratio of important words for positive versus negative trading days

Sentiment Analysis

Advanced strategy

- Utilize NLTK package
- Group tweets by date but treat each tweet independently
- Pair natural language processing (NLP) and machine learning algorithms together to aggregate and measure individual emotions during the trading day.
- Detect market emotion that can be used as signal for trade recommendation

Data Sources



Overall Data Flow

Data Inputs

BTC price

- Historical (.csv)
- Live (API/DB)

Tweet Data

- Historical (.csv)
- Streaming (DB)

Trading Algorithm and Performance Evaluation

Python Script

- Extract / Load
  - Bitcoin pricing
  - Tweet data
- Transform data
- Generate trading signals
- Track trades and portfolio performance

Export

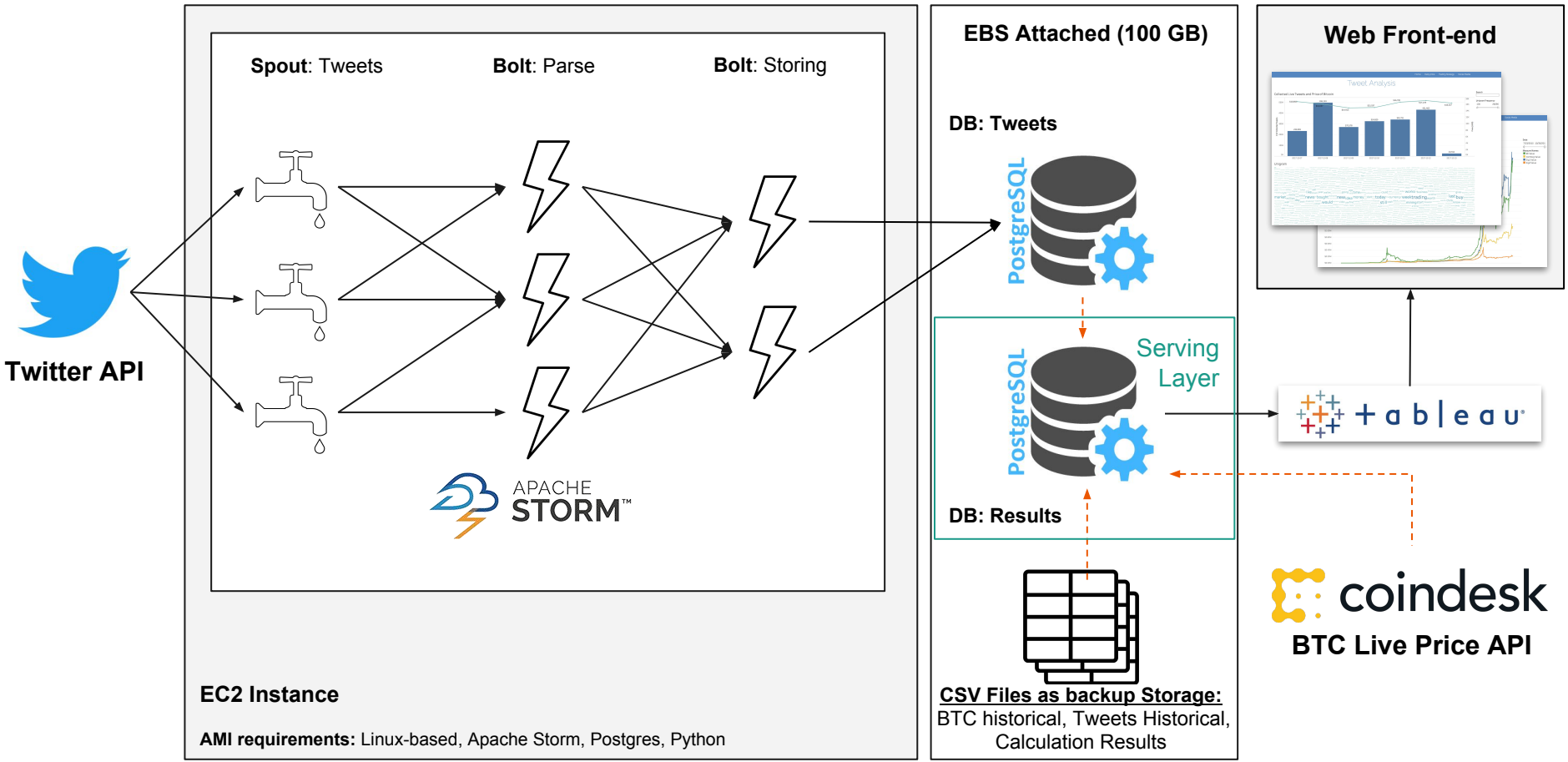
Write results to postgres DB for:

- Ongoing performance tracking
- Visualization
- Algorithm refinement

Data & Functional Requirements

Area	Functional Requirements	Specifications
Data Lake / Datamart	• Storage of Twitter and BTC historical data (<2MB) • Streaming of Twitter and BTC data (2M tweets, 10 GB/week)	• EC2 m3.xlarge + 100 GB Attached Storage (limited spec for prototype) • Apache Storm, write to PostgresSQL
Staging & Processing	• Parse & Aggregate Twitter Data • Apply Algorithm for Trading Signals	• Python 2.7+ for Data Processing & Algorithm Scripts, 7 day batch
Serving Layer	• Storage of processed data in tables for querying	• PostgreSQL, Schema on Write
Data Visualization	• Merge Tables & Create Views • Visualize & Analyze Trading Performance and Trends	• Tableau Workbooks • HTML iFrame for Live Dashboard

High-level Architecture



Live Trading Dashboard

[Go to Live Dashboard >]

Insights

- **Trading Findings from Historical Analysis**
  - 5 Day Moving Average Performed Best (before transaction costs and taxes) as it limited portfolio drawdowns during steep price drops
  - Tweet signal was the worst performing strategy,
    - Ineffectiveness could be due to lower tweet volume and that using tweet counts fails to extract the information contained within tweet text
    - **Note**: Twitter grew 30M to 330M monthly active users between 2010 to 2017, Bitcoin related tweet volume grew from ~300/day to ~350k/day (+1,167x)
  - Other considerations for future implementation:
    - Optimize lookback window
    - Optimize signal weights (currently 50/50)

Historical Performance Evaluation

```
Strategy Performance Summary

Value of initial $1,000 investment at end of performance measurement period (07/23/2010 - 10/30/2013):

Buy and Hold: $3,166,554
5DMA: $3,262,248
Tweet 5DMA: $221,003
5DMA + Tweet Strategy: $1,134,943


Risk Adjusted Returns - Trading Strategy Sharpe Ratios

Buy and Hold: 0.13
5DMA: 0.15
Tweet 5DMA: 0.11
5DMA + Tweet Strategy: 0.14


Trade Summary

Total Number of Trading Days: 1,196

Total Number of Trades Executed: 521

No Trade              675
REDUCE to 50%         142
ADD to 100%           126
REDUCE to 0%          115
ADD to 50%             99
BUY FULL POSITION      28
SELL FULL POSITION     11
Name: trade_type, dtype: int64
```

Insights

- **Scaling up:**
  - Architecture is flexible - even though data is collected daily, we can collect hourly or even to the minute with a few lines of code changed. With a cost factor.
  - Storage Cost: 10gb (2M tweet) in one week. ~520 GB/year. ~$1,000/year.
  - EC2 Instance Cost: ~$50/week. ~$2,600/year
  - Implement Zoomdata to deliver streaming visualization output

- **Challenges:**
  - Limited credit did not allow us to scale to the infrastructure requirements

Helpful links

- http://www.livebitcoinnews.com/can-twitter-sentimental-analysis-predict-bitcoin-price-fluctuation/
- https://medium.com/@SamuelCouch/understanding-cryptocurrencies-with-sentiment-analysis-5fc4cf66ec28

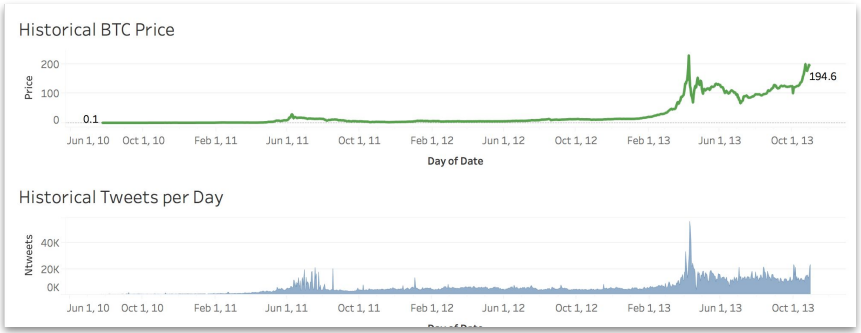
Evaluating Trading Algorithm Effectiveness

- Compare the performance of paper portfolios implementing the following strategies:
  - Buy and Hold
  - Base Price Momentum Signal
  - Tweet Signal
  - Combined Price Momentum Signal + Tweet Signal

- Performance Evaluation Caveats
  - Execution Costs
  - Trading Commissions
  - Taxes

Tableau Dashboard

Historical BTC Price vs. Number of Tweets



Historical Data Training Algorithm