**EECS E6893: Big Data Analytics Final Project Proposal**

Cheng-Hao Ho(ch3561) Shuoting Kao(sk4920) Wei-Ren Lai(wl2777)

**Goal — novel? Challenging?**

The motivation is to make a lot of money and provide a powerful tool for investors to increase their profit. Our platform predicts the crypto’s trend based on cryptocurrency historical transaction data, wallet transaction data, and sentiment analysis from the various data sources, including Twitter, Crypto News, Reddit, and Google Trend. Besides, our goal is to trace top performers’ activities and detect the key signal through their activities to decide whether we should follow their moves.

* **Novelty**

1. Predict the trend of crypto based on various data types and sources - price, volume, wallet data, transaction data and text data
2. Monitor large trader’s activities and decide if we can follow them based on ML/DL models or data mining algorithms
3. Detect important events or signals that relate to the price fluctuation and notify the users
4. Event subscription
5. Take model confidence level into account
6. Novel ways to make money and help people retire early

* **Challenges**

1. How to classify the source data is a challenge
2. Need to scrape data by ourselves
3. Learn how to use various APIs
4. The result should be real-time
5. Prediction accuracy
6. Data storage should be a big challenge due to the tremendous amount of transaction data on the blockchain

**Data — Big Data 3Vs? New or Existing dataset (from where)?**

We build a new dataset by ourselves by scraping data from Twitter, Reddit, Google and Crypto news API.

* **Volume**

1. The volume of transaction price/volume
2. Twitter posts
3. Reddit posts
4. Crypto news
5. Google trend

* **Velocity**

1. Twitter streaming data
2. Transaction streaming data

* **Variety**

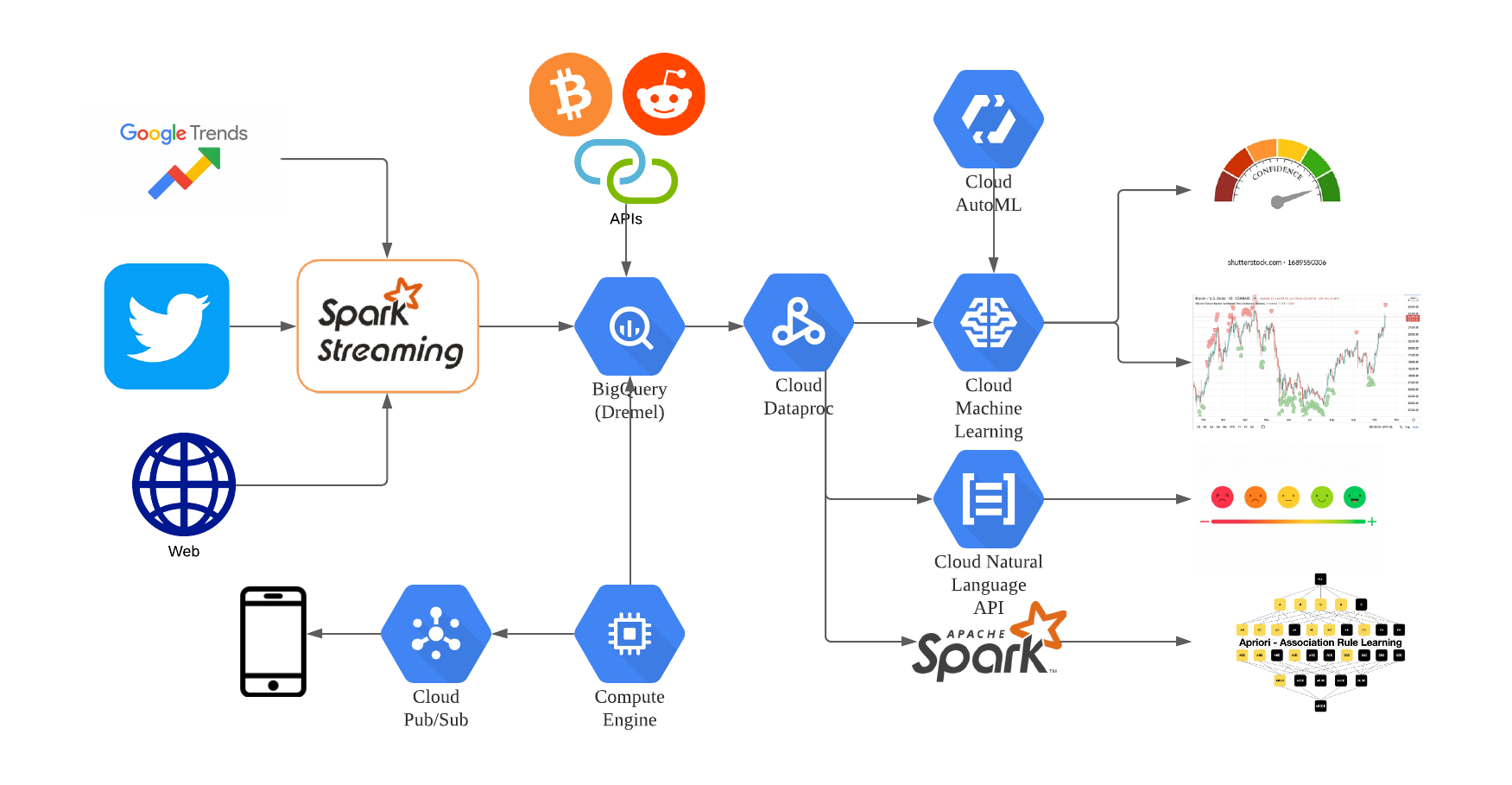
1. Cryptocurrency transaction structural data
2. Cryptocurrency wallet in and out data
3. Twitter text data
4. Reddit text data
5. Crypto news data
6. Google search popularity

**Methods — potential methodologies and algorithms? Feasible?**

1. Time Series Analysis
2. Association Rules Algorithm (PrefixSpan)
3. Sentiment Classification
4. Machine Learning & Deep Learning

**System — an overview of system. Multiple steps? Scalable?**

1. Data acquisition. Two streaming programs. One processes twitter posts and the other one handles the transaction
2. Use self-built dataset to train machine learning / deep learning model
3. Deploy a model endpoint for real-time prediction
4. Automate the model re-training
5. Event monitor addresses large traders’ activities and send notification to subscribers



**Schedule — what to achieve by what time, and by whom?**

1. 11/19 progress slide (11/19 ~ 12/3)
   1. Data Collection from Web and association rules mining by Wei-Ren
   2. Build machine learning/deep learning model by Cheng-Hao
   3. Streaming data analysis & sentiment analysis by Tim
2. 12/3 progress report (12/3 ~ 12/17)
   1. Finish all preliminary model prediction by the selected model
   2. Try to combine our model prediction data with Twitter posts / Reddit posts / Crypto news
   3. Predict the future of the Crypto price and judge whether our model is reliable
   4. Visualize the model prediction
   5. Predict the model confidence level and check its accuracy.
3. 12/17 project final demo  
    a. Demonstrate full features of the service.

b. Document the design and implementation.

Reference

<https://medium.com/general_knowledge/watch-the-whales-101-guide-to-wallet-tracking-8ff5799f3dc4>

<https://coinyuppie.com/how-to-track-whales101-wallet-tracking-guide/>

<https://rapidapi.com/collection/best-bitcoin-apis>

Crypto news API

<https://polygon.io/crypto?gclid=CjwKCAjwoP6LBhBlEiwAvCcthFZ4jA5U9iOCeMSbxOdTwiAYWXkSgAPBFYcmwM3NRFg3CnqjD5YWHhoCk78QAvD_BwE>

Slides

<https://docs.google.com/presentation/d/16NiqikiaOTEPcJ_7Bk-TiUwFNV1zMmcICJ20qgRwag0/edit#slide=id.gfd220f5eb6_0_8>

<https://docs.google.com/spreadsheets/d/1YwQ_ZBrpRDCOThySZ_uLwd3mlvxlPJjUc356d_rf2WU/edit#gid=0>

<https://lucid.app/lucidchart/3f13d52a-fde4-4e37-9678-722149f87f02/edit?beaconFlowId=0C0EACCF52AE0E8C&invitationId=inv_25c2d13a-c453-4e15-958b-f9d1e30861c4&page=0_0#>