Nowcasting Food Prices using Social Media

Jong Gun Lee

Data Scientist
UN Global Pulse

'NOWCASTING' COMMODITY PRICES IN INDONESIA

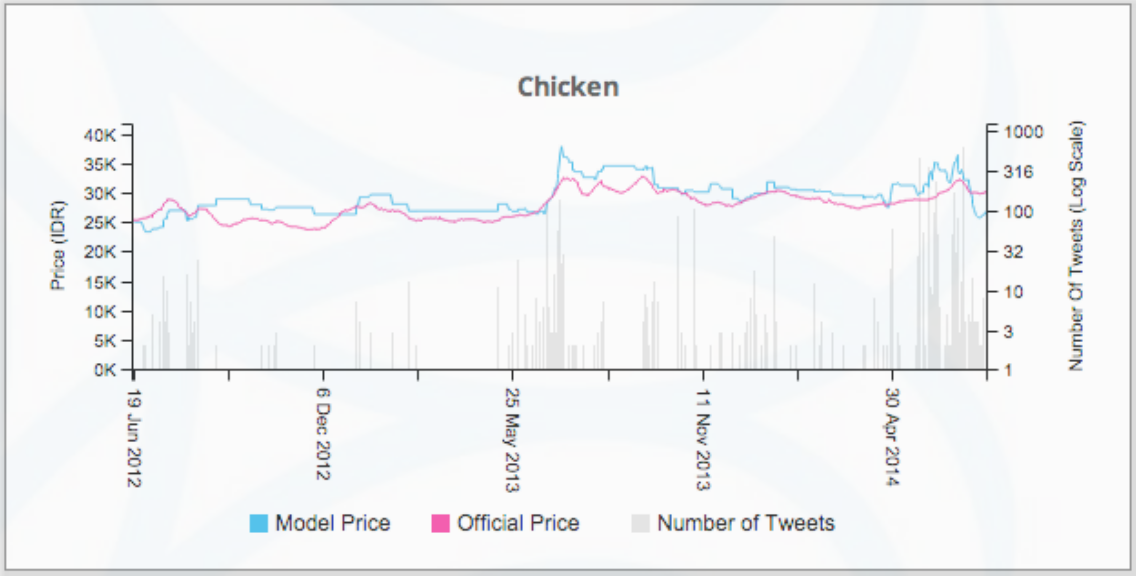
Can estimates of food prices be extracted from social media?

How do these estimates compare to official statistics?

Pulse Lab Jakarta conducted a research project to investigate the feasibility of 'nowcasting' commodity prices in Indonesia, a country famously addicted to social media. This page presents near to real-time price information on three staple food commodities: beef, chicken and onion, using tweets about food prices in Indonesian language Bahasa Indonesia.

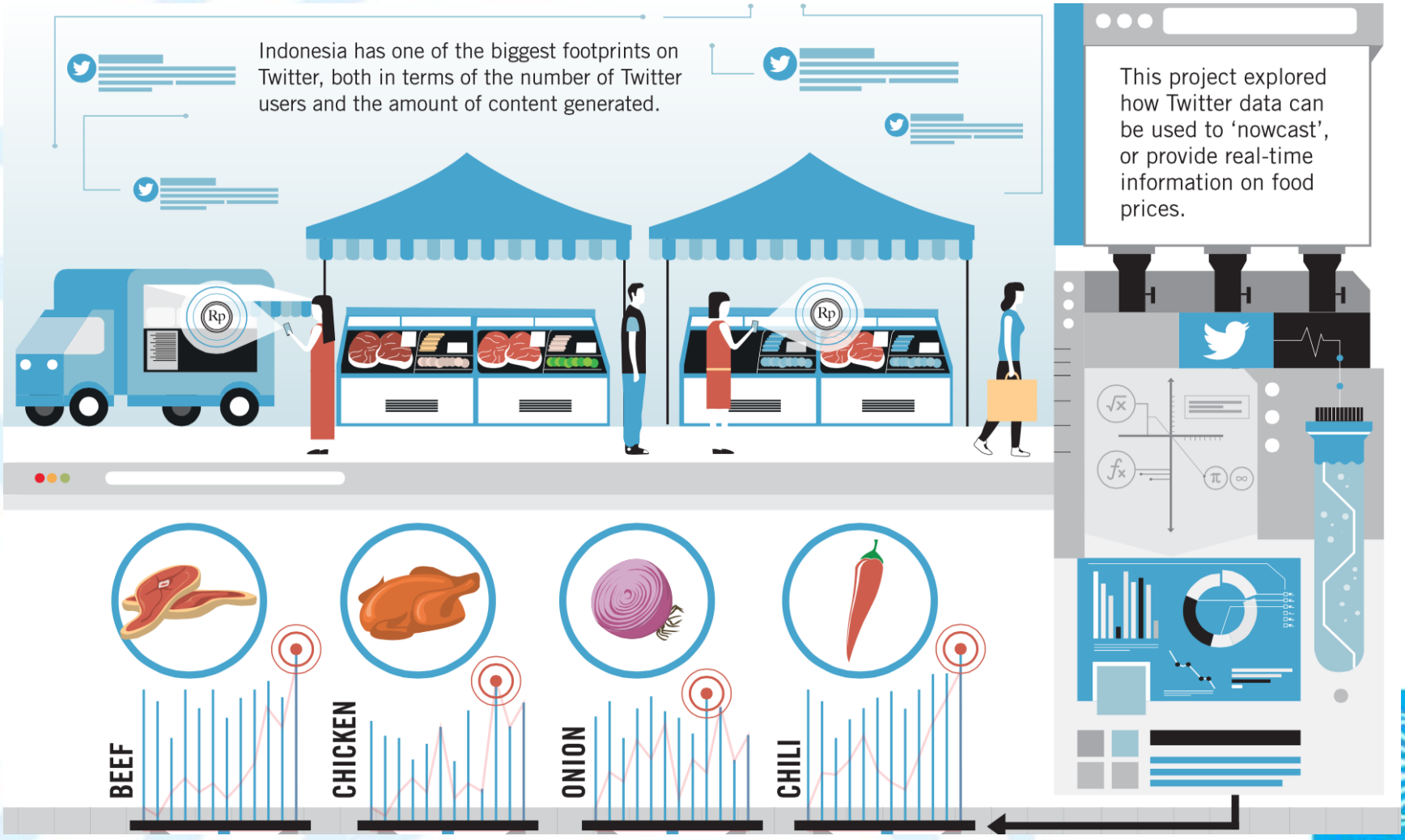
This project has been done in collaboration with the Indonesian Ministry of National Planning Development (Bappenas) and the World Food Programme Indonesia.

For more information on this project, please visit: http://www.unglobalpulse.org/nowcasting-food-prices



GLOBAL PULSE

Nowcasting Food Prices in Indonesia using Social Media Signals



http://unglobalpulse.org/nowcasting-food-prices

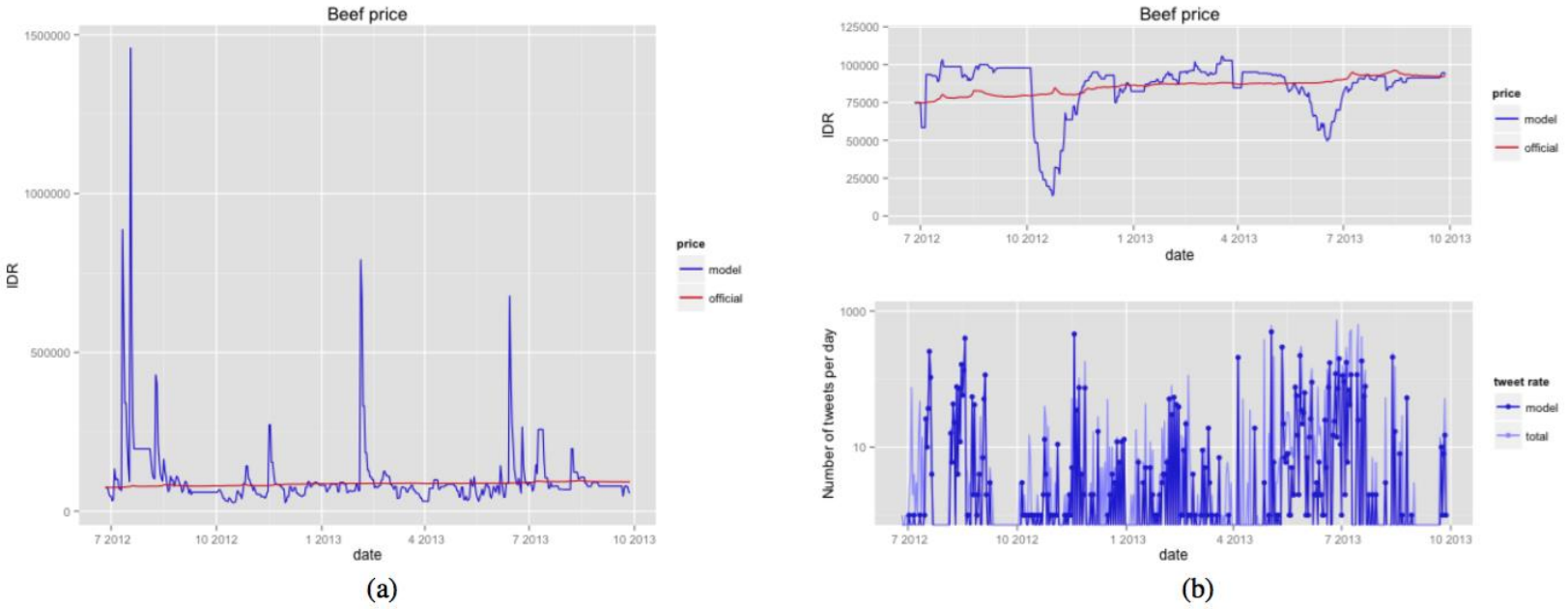


Figure 9. (a) Model running test with entire raw dataset for beef (14473 tweets). Outlier values made some points with huge spike from ground-truth price. Mean Absolute Errpr (MAE) is 43814.22 (b) Model running test with IQR filter (7908 from 14473 tweets picked up after filtering). MAE is 11911.59

A Generative Model

Hypothesis - "Today's price $P_{(i+1)}$ is inferable from yesterday's price $P_{(i)}$ and today's tweet price $P_{tweet(i+1)}$"

$$P_{i+1} = \frac{\alpha P_i + \beta P_{i+1}^{tweet}}{\alpha + \beta}$$

GLOBAL PULSE

A Generative Model

Hypothesis - "Today's price $P_{(i+1)}$ is inferable from yesterday's price $P_{(i)}$ and today's tweet price $P_{tweet(i+1)}$"

$$P_{i+1} = \frac{\alpha P_i + \beta P_{i+1}^{tweet}}{\alpha + \beta}$$

$$P_{i+1}^{tweets} = \frac{\sum_{j=1}^{[T_{i+1}]} w_{i+1}^j T_{i+1}^j}{\sum_j w_{i+1}^j}$$

$$w_{i+1}^j = \begin{cases} 1 - \frac{\left| \frac{T_{i+1}^j - P_i}{P_i} \right|}{\delta} & ,\text{if } \left| \frac{T_{i+1}^j - P_i}{P_i} \right| \leq \delta \\ 0 & ,otherwise \end{cases}$$

$$P_i = \frac{\sum_{j=i-k}^{i-1} P_j}{k} \text{ where no tweets over } n \text{ days}$$

Components for weighing and filtering as well as refreshing

GLOBAL PULSE

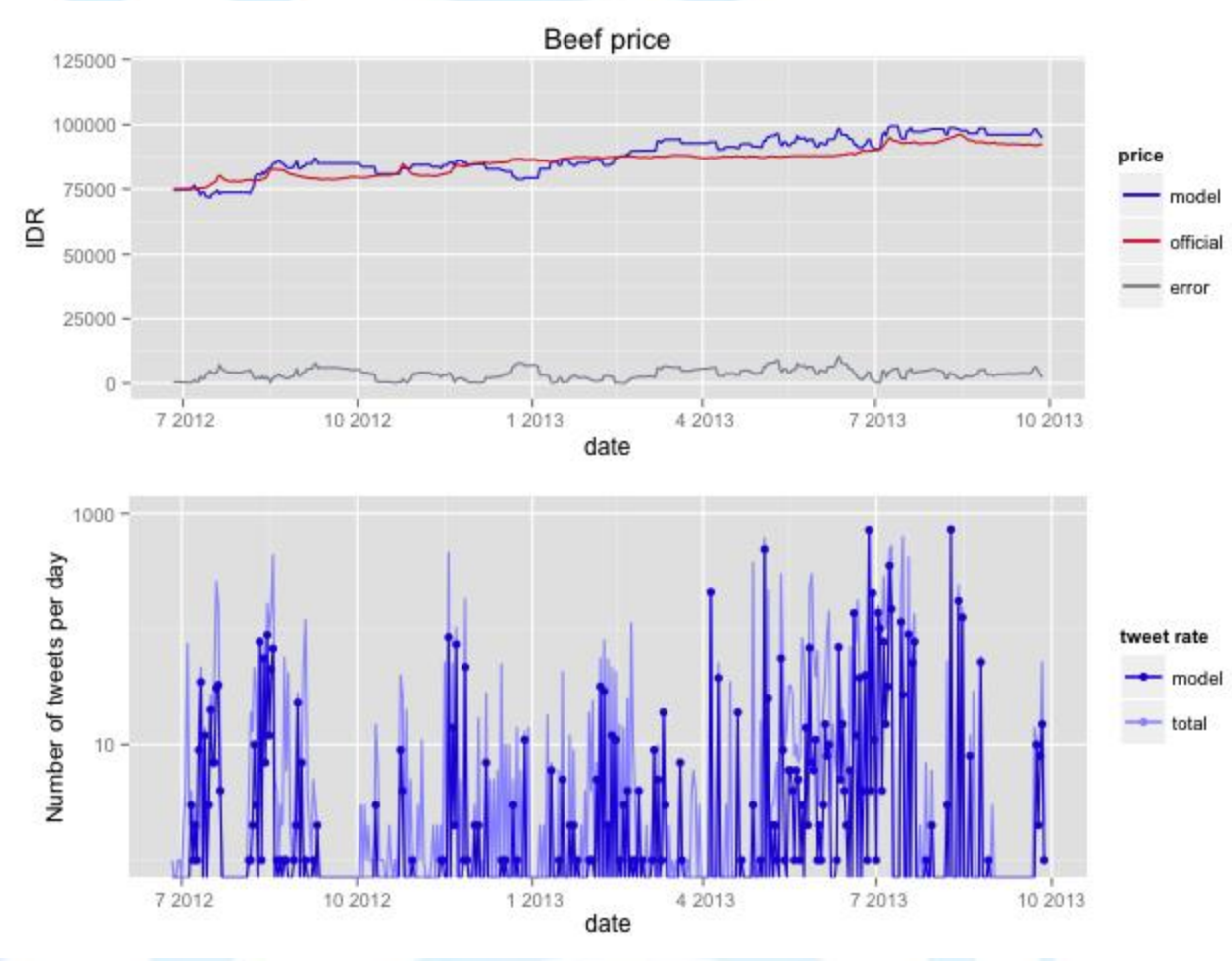
Data

- Primary data
  - About 23,000 tweets for the following four commodities
    - Beef, Chicken, Chili, and Onion
  - Taxonomy
    - Commodity name & Price (number & currency)
    - Unit, if available
- Secondary data
  - Daily price information from Indonesian Ministry of Trade
- Duration
  - 16 months between 2012/07 and 2013/10

GLOBAL PULSE

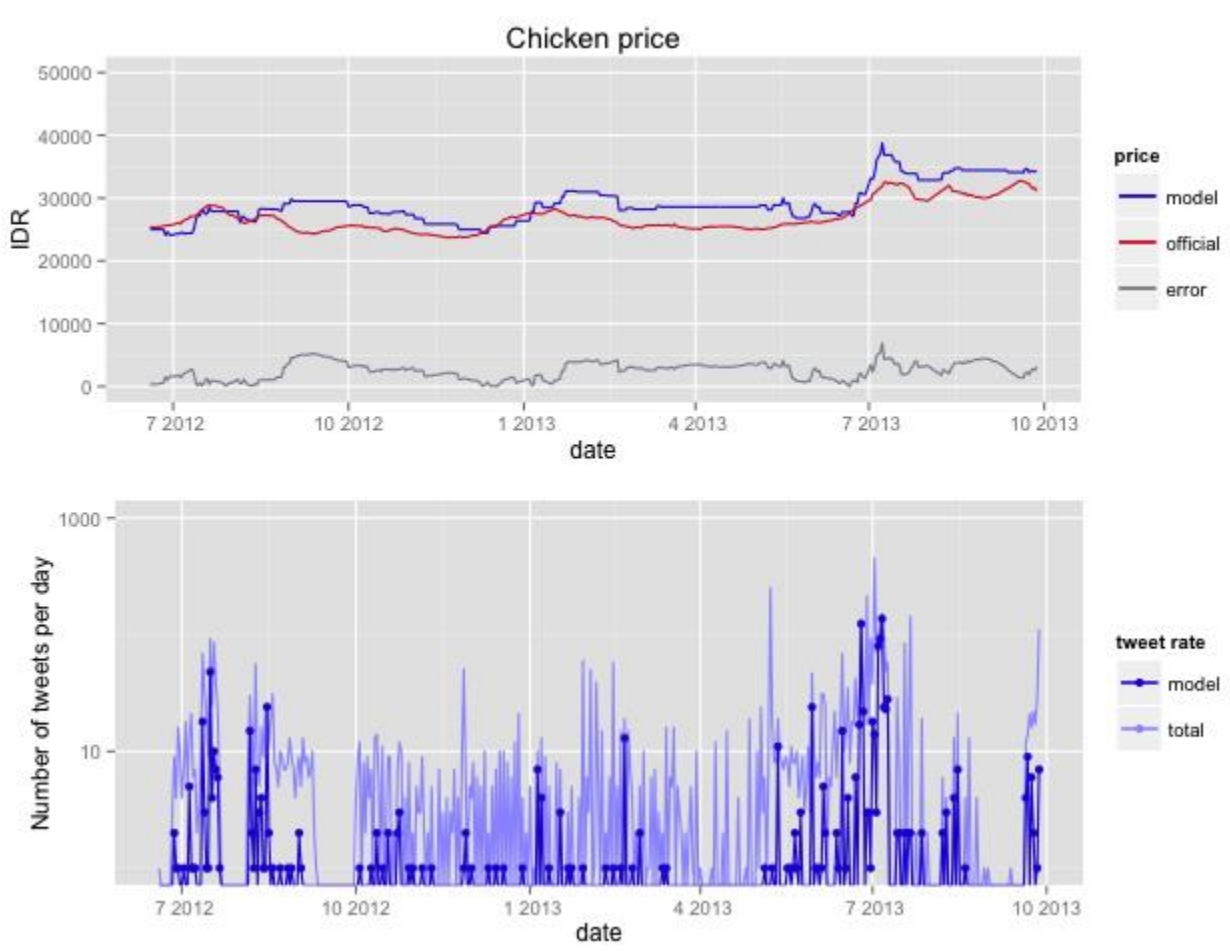
Beef (daging sapi)

- 5,707 tweets used among 14,473 tweets
- Pearson correlation - 0.87



Chicken (daging ayam)

- 974 tweets used among 5,223 tweets
- Pearson correlation - 0.81



Onion (bawang)

- 610 tweets used among 1,954 tweets
- Pearson correlation - 0.85

Chili (cabe)

- 206 tweets used among 1,772 tweets
- Pearson correlation - 0.76

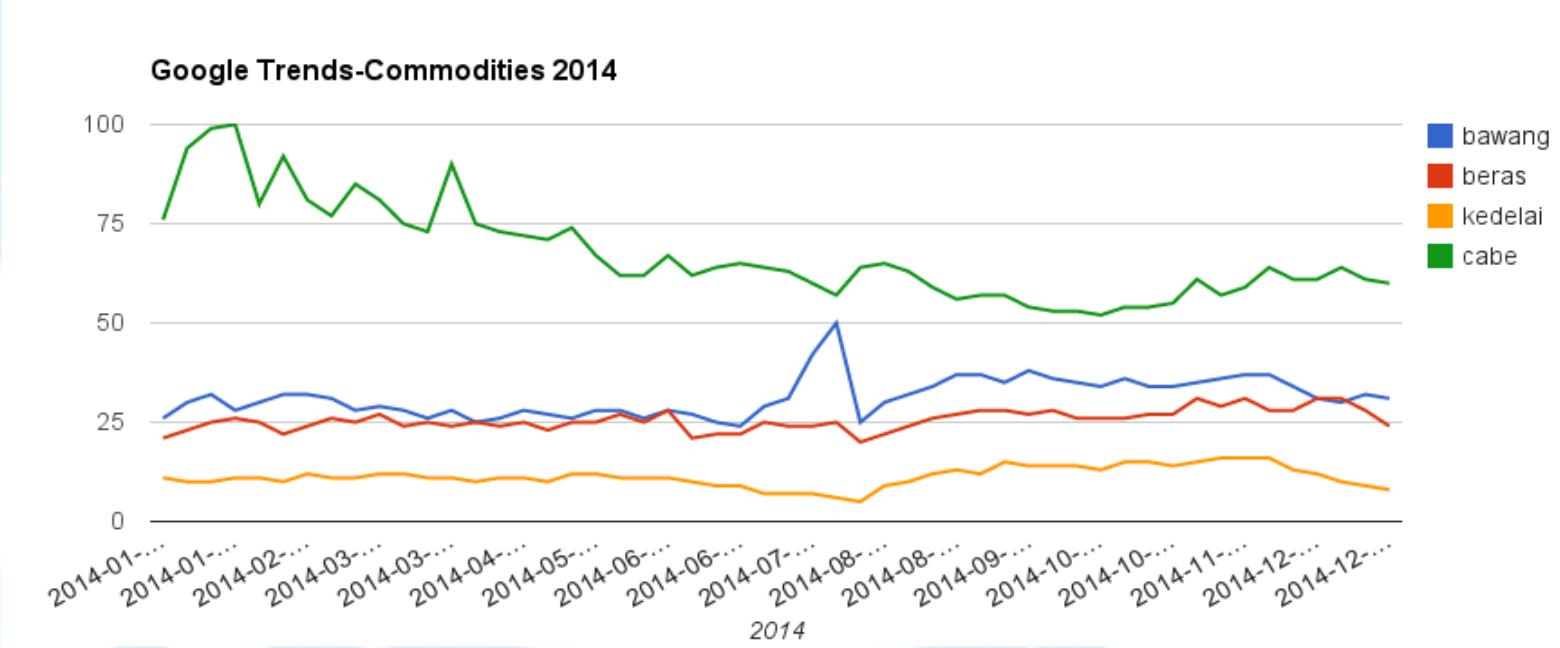
GLOBAL PULSE

NOWCASTING FOOD PRICES (ONGOING)

GLOBAL PULSE

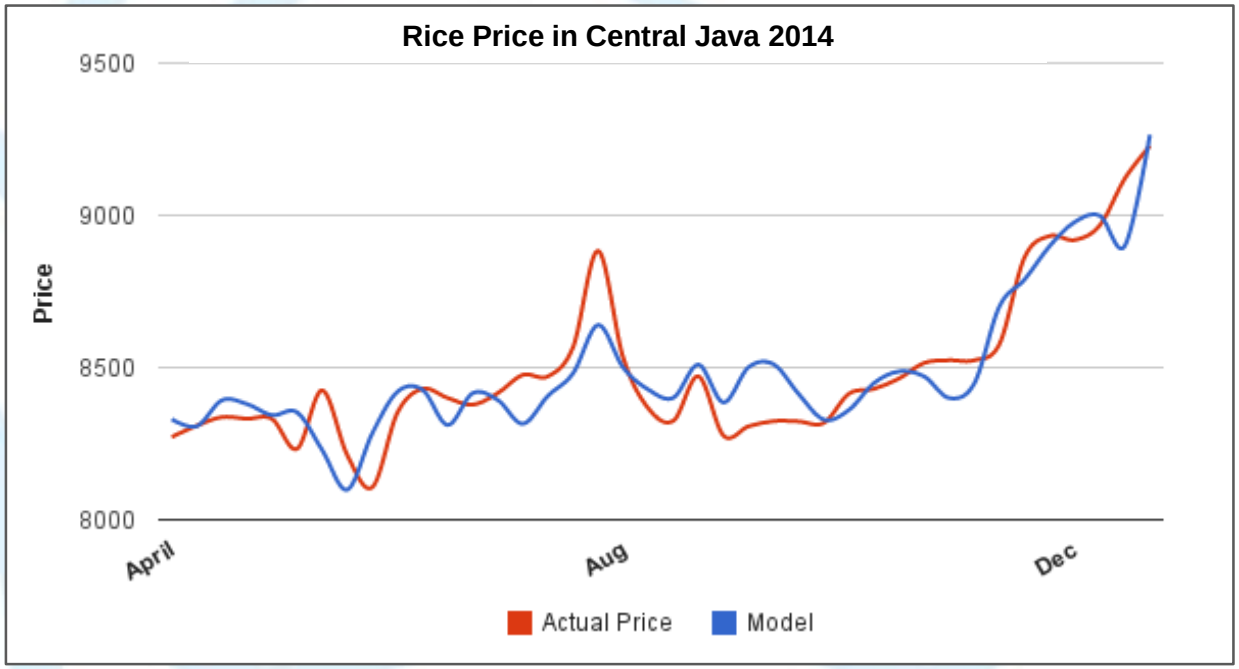
Google Trends Data

- Google Trends service provides normalized, aggregated search volume per day per keyword up to provincial level.



A Simple Linear Model

"Provincial Price = Past National Price + Recent Google Trends in Province "



$$P_i = \alpha + \beta_1 MoT_{i-2} + \beta_2 MoT_{i-4} + \beta_3 GT_i + \beta_4 GT_{i-1} + \beta_5 GT_{i-3}$$

MoT_{i-1}: National price two weeks ago

MoT_{i-4}: National price four weeks ago

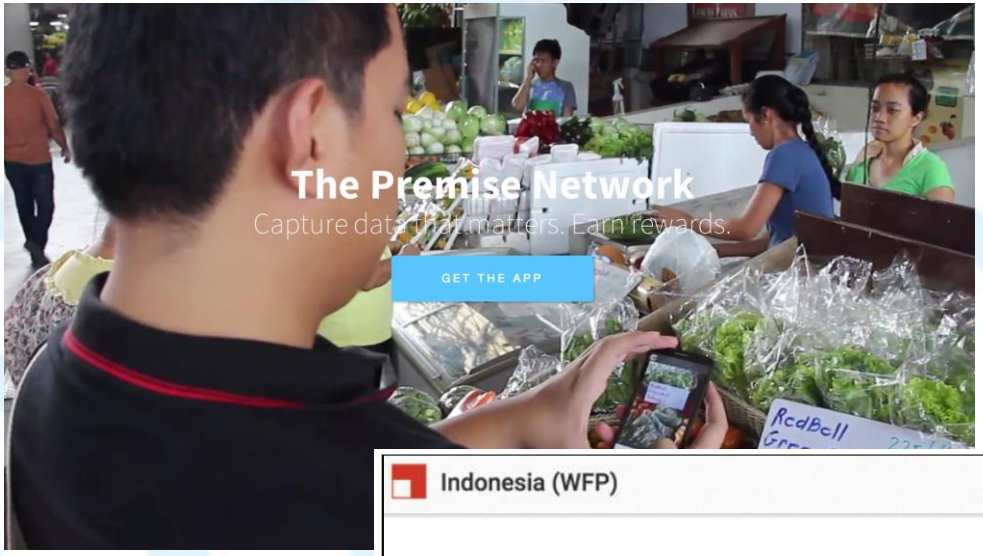
GT_i: Google Trends this week

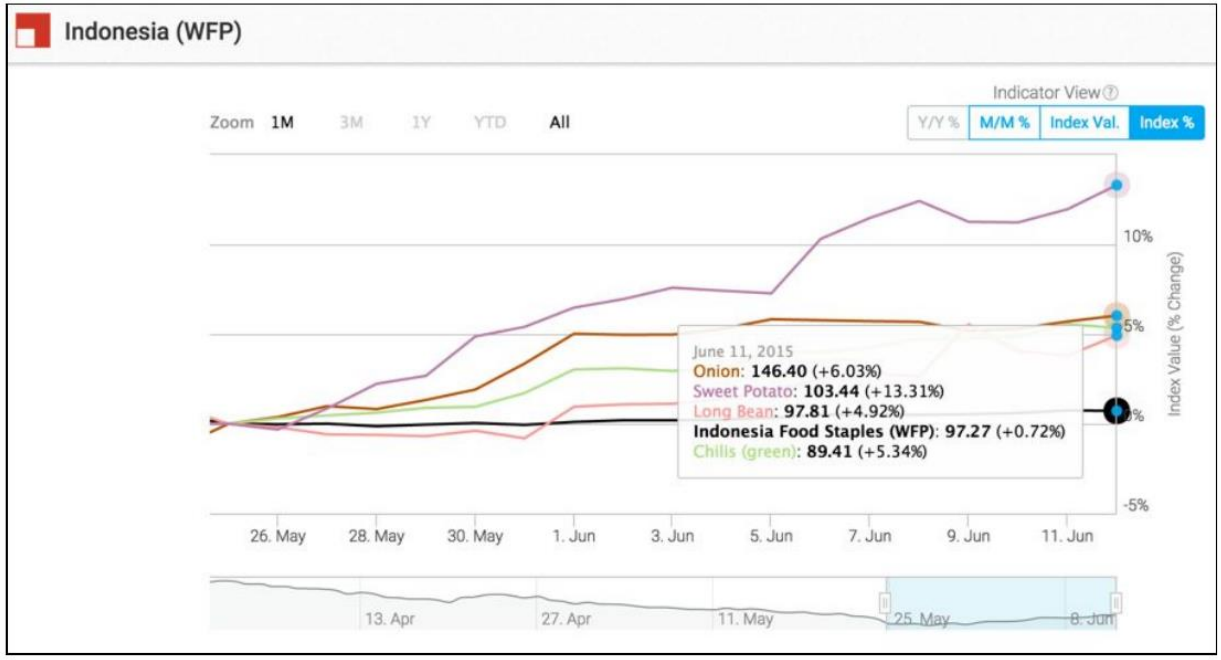
GT_{i-1}: Google Trends last week

GT_{i-3}: Google Trends three weeks ago

GLOBAL PULSE

New Digital Data Sources + Reports by Citizens?





GLOBAL PULSE

jonggun.lee@un.or.id
www.unglobalpulse.org

@unglobalpulse
@pulselabjakarta
@pulselabkampala