



NOWCASTING FOOD PRICES IN INDONESIA USING SOCIAL MEDIA SIGNALS

PARTNERS: UN WORLD FOOD PROGRAMME, INDONESIA MINISTRY OF DEVELOPMENT PLANNING (BAPPENAS), KOREA ADVANCED INSTITUTE OF SCIENCE & TECHNOLOGY
PROGRAMME AREA: FOOD & AGRICULTURE



SUMMARY

This project explored how Twitter data can be used to ‘nowcast’ or provide real-time food prices. Pulse Lab Jakarta collaborated with Bappenas and WFP on the study, the outcome of which was a statistical model of daily price indicators for four food commodities: beef, chicken, onion and chili. When the modeled prices were compared with official food prices, the figures were closely correlated, demonstrating that near real-time social media signals can function as a proxy for daily food price statistics. This preliminary study paves the way for further research on how social media analytics can complement traditional price data collection by offering a faster, more affordable and efficient way of collecting real-time food prices.

BACKGROUND

The ability to monitor market commodity prices in real-time could provide critical information for policy decisions on food security and other economic issues. However, not all countries have well-established systems for accessing and monitoring price data, especially at a high frequency. New approaches in price monitoring could help to fill this data gap in developing economies.

Pulse Lab Jakarta conducted a study to explore Twitter as a data source for building a real-time food price tracking model in Indonesia. Indonesia was chosen as the case study country for two primary reasons. First, official commodity price data is available on a daily or weekly basis, as the Indonesian government already uses a dependable monitoring system for commodity prices. The Ministry of Trade collects and publishes daily price information, and the Indonesian Bureau of Statistics (BPS) collects weekly commodity price information. Second, Indonesia is one of most-tweeting countries, both in terms of number of Twitter users and amount of generated content.

USING TWEETS TO MODEL FOOD PRICES

To refine the scope of the research study, four food commodities (beef, chicken, onion and chili) were selected to track. These commodities were chosen based on data availability and the country-level priorities for food security monitoring in consultation with the Ministry of National Development Planning and World Food Programme (WFP) Indonesia. Data was analysed from June 2012 to September 2013, a timeframe of 15 months.

In order to extract prices as quoted on social media, a taxonomy of keywords and phrases related to the four food commodities was used to extract relevant public tweets, which were filtered further for those tweets containing price quotes. Here is an example tweet in both Bahasa and English that mentions a food price:

Pemerintah lagi sibuk! RT @promoasyik: Capai Rp 100 Ribu per Kg, Harga Cabai Rawit Merah Setara Daging <http://bit.ly/1pUmUmo>

The government was busy! RT @promoasyik: Reaching Rp 100 thousand per kg, red chili is equivalent to the price of meat <http://bit.ly/1pUmUmo>

A numerical filter was then used such to ensure the tweet price did not exceed a maximum allowable daily percentage price change (itself based on historical rates). This filter was validated through an examination of the historical series of official food prices (for example, a daily change of 3.5% was the largest change observed in the beef price dataset). The numerical model was also adapted to consider both the daily number of tweets and the divergence of prices reported in tweets from the last official price.

Using the keyword and numerical filters to extract 41,761 relevant tweets, a statistical model was then built in which the proxy for daily commodity price was based on data including:

- Model price from the day before, as quoted on Twitter
- Volume of tweets (higher volumes of price quotes would make these quotes more credible)
- Maximum daily price change rate

This nowcasting model was compared with official food prices, aggregated daily across Indonesia. ‘Ground truth’ price data provided by the Ministry of Trade was aggregated across the entire country on a daily basis.

INSIGHTS & OUTCOMES

The model for nowcasting commodity prices gave high correlation with official food price statistics over the timeframe covered.

The results of the model are shown in the table below:

COMMODITY	PEARSON CORRELATION COEFFICIENT
Beef	0.87
Chicken	0.81
Onion	0.85
Chili	0.76

The research findings led to development of a near real-time food price index to extract daily food prices from public tweets.

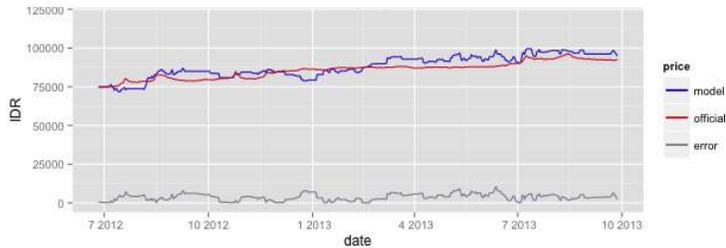
HOW TO CITE THIS DOCUMENT:

UN Global Pulse, ‘Nowcasting Food prices in Indonesia using Social Media Signals’, Global Pulse Project Series no. 1, 2014.

The following four graphs show the official commodity price (in red) compared with the statistical nowcasted price (in blue) from July 2012 to October 2013.

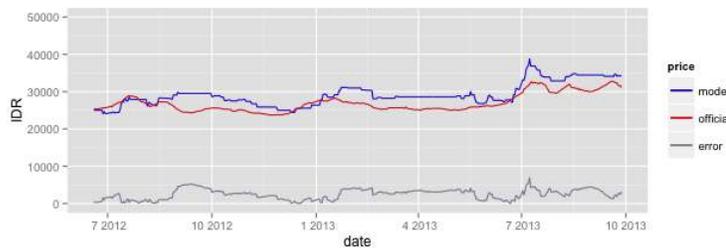
BEEF: The official beef price was relatively stable for the timeframe analysed, and the long-term price trend was accurately modeled (nowcasted prices remained close to real prices throughout the timeframe).

14,473 tweets contained price quotes on beef.



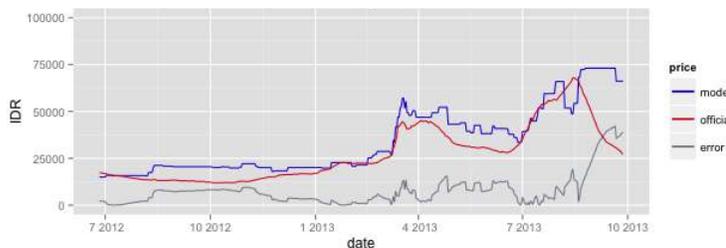
CHICKEN: The tool successfully modeled the price of chicken (“daging ayam”) with significant correlation for the duration.

5,223 tweets contained price quotes on chicken.



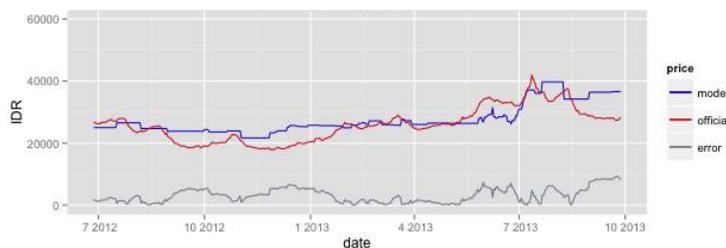
ONION: Onion (“bawang”) fluctuated the most of the four commodities. The model effectively tracked price variation around April 2013, but due to the low number of tweets containing price quotes in August did not provide a successful price proxy.

1,954 tweets contained price quotes on onions.



CHILI: Chili (“daging sapi”) showed significant price correlation, although the proportion of relevant tweet data was much lower than the other three commodities analysed.

1,772 tweets contained price quotes on chili.



CONCLUSIONS

This study showed that Indonesians do tweet about food prices, and that those prices closely approximate official figures. A near real-time food price index, with prices nowcasted using social media signals, is an efficient tool with immediate utility for policymaking and economic risk management.

- The study proved the potential accuracy of nowcasting by comparing prices extracted from public tweets with official market prices.
- This tool could provide early warning for unexpected price spikes at much lower costs than traditional data collection.
- The simple and replicable technical needs and methodology—keyword taxonomy refined by numerical filters—allows for straightforward operational implementation and scaling.

The nowcasting model could potentially be adapted to other types of commodities beyond the four food commodities analysed, or used in other languages and countries. More complex analysis methodologies to extract the information from the messages could be developed (eg. based on natural language processing) to improve the accuracy of the model. Finer-grained official price statistics would be necessary to conduct and validate a more detailed analysis (for example, by analysing price by location using social media geolocation techniques).

IMPLICATIONS & RECOMMENDATIONS

- This research shows the potential for nowcasting food commodity prices by extracting daily food prices from public tweets to generate a near real-time food price index.
- The model prices correlated closely with official food price statistics over the timeframe covered, demonstrating the potential for social media signals to be used as a proxy to model food prices.
- It is recommended this data mining approach is further validated and applied to other commodity prices, in new locations and using additional social data sources.
- It is recommended to develop a framework for potential integration of near real-time indicators derived from social data into existing national statistics agency programmes.