



BADAN PUSAT STATISTIK



Big Data Implementation for Price Statistics in Indonesia

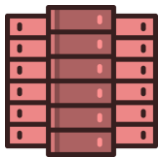
Setia Pramana

Politeknik Statistika STIS
Badan Pusat Statistik

Price Statistics



- Consumer Price Index (CPI) is one of the important economic indicators that can provide information about prices of commodities paid by consumers
- Currently, most of National Statistics Offices (NSO), including, BPS-Statistics Indonesia, rely on collecting commodities' prices manually through a direct field survey
- Advance Big Data technologies and vast data sources has National Statistical Offices to experiment in creating price indexes based on new alternative sources e.g., scanner data and web scrapping data from online shops or marketplaces.
- Scanner data are digital transaction data on sales, price and type of items sold recorded at the retail shops. Web scraping extracts data from any publicly available information of websites.



VOLUME

Scale of Data

- The volume of persistent usable data in analytics system at any point in time.

Click Stream	Printed Corpus
Active/Passive Sensor	Speech
Log	Social Media
Event	Traditional



VARIETY

Forms of Data

- The form and content of data structured (RDBMS), semi structured (Social Media) or unstructured (text/documents).

Structured	Unstructured
	Semi-structured

BIG DATA



VELOCITY

Analysis of Data-Flow

- How quickly the analytics system process the data to create insights.

Speed of Generation	Rate of Analysis
---------------------	------------------



VERACITY

Uncertainty of Data

- The degree to which data is accurate, precise and trusted.

Uncleansed	Untrusted
------------	-----------



- A term -> describe extremely large amounts of structured and unstructured data.
- The activity -> capture/storage/processing/sharing/reporting of data -> beyond ability of legacy software tools and hardware infrastructure.
- Related to many "science" branch -> data analytics, data science, machine learning, artificial intelligence, IoT, and many more.
- The application -> on many field -> efficient, cost effective, faster and accurate decision making.

TAXONOMY BIG DATA SOURCES



EXHAUST DATA



Passively collected data from people's use of digital services such as mobile phones, financial transactions or web searches.



SENSING DATA



Actively collected data from sensors, e.g. in smart cities or from wearables and also through remote sensing and satellite images.



DIGITAL CONTENT









Open web content actively produced by people such as social media interactions, news articles, blogs or job postings. Unlike exhaust and sensing data is digital content intentionally edited by somebody, i.e. subjective or even deceptive, depending on the intentions of the author.

Big Data Initiatives and Developments in BPS

Web-crawling



 Marketplace	▶	E-commerce Data
 Flight Tracker, bus booking site	▶	Transportation analytics
 Job Vacancy Site	▶	Labor analysis
 Online booking site and review	▶	Room occupancy rate, Number of tourists, et
 Air Quality, weather reporting site	▶	Enviromental and disaster statistics
 Mobil123, rumah 123	▶	Property and vehicles statistics
 Online news and social media	▶	Current fenomena, citizen sensing
 Google Map	▶	Infrastructures and people activities
 IDX	▶	Company financial report, Stock index

Google and Facebook mobility index



People mobility

Satelite Imagery



Economic activities, Agriculture statistics
Poverty mapping

FB Relative Wealth Index



Economic activities
Poverty mapping

Mobile Positioning Data



Tourism statistics,
Metropolitan Statistics Area

Utilization of Big Data for Price Statistics

Australia	scanner data from supermarket	official	2011	https://www.abs.gov.au/statistics/research/recent-applications-supermarket-scanner-data-national-accounts
China	e-commerce data, scrapping, transaction data from supermarket, partnership with companies to obtain original price data	official	NA	https://unstats.un.org/bigdata/events/2019/hangzhou/presentations/day1/3-1.%20Progress%20of%20Big%20Data%20for%20Official%20Statistics%20in%20China.pdf
Italy	scrapped data	pilot project	2015	https://www.igi-global.com/chapter/big-data-techniques-for-supporting-official-statistics/217860
Germany	scrapped data	official	2019	https://www.destatis.de/EN/Themes/Economy/Prices/Consumer-Price-Index/background-cpi-revision.pdf?__blob=publication;Filehttps://www.destatis.de/EN/Methods/Quality/QualityReports/Prices/consumer-price-index.pdf?__blob=publicationFile
New Zealand	scanner data	official	2014	https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/Measuring-price-change-for-consumer-electronics-using-scanner-data/measure-price-change-electronics-scanner-data-2014.pdf



Using Big Data to Nowcast Food Prices



Collaboration with Pulse Lab UN Jakarta

- ❖ *Aim: to nowcast food prices using multiple sources of data including social media, Google Trends, and crowdsourcing as well as official statistics from BPS, Ministry of Trade and Ministry of Agriculture.*
- ❖ *Locus: Kota Mataram, NTB*
- ❖ *Time: March–July 2015*

DATA SOURCES

CROWDSOURCE

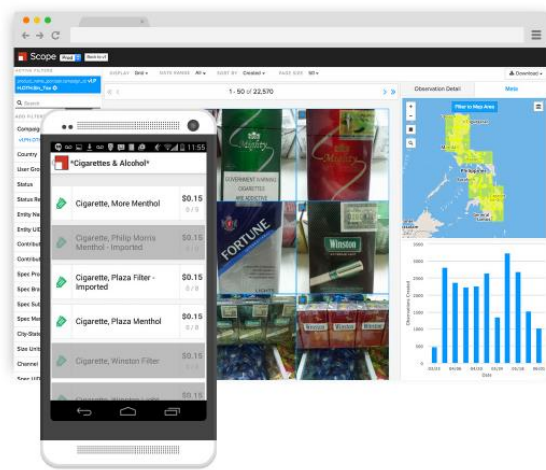


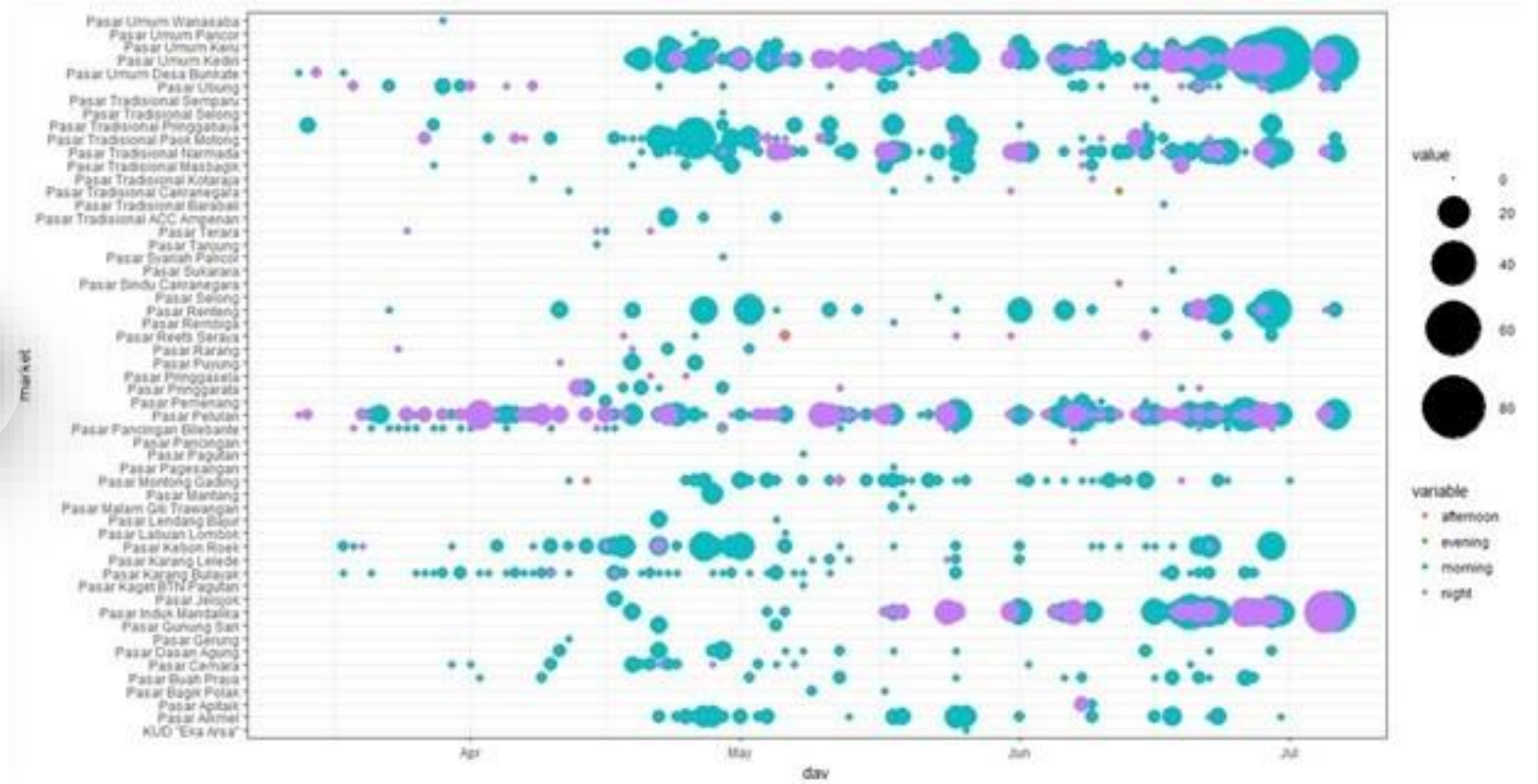
Figure 14: The Premise mobile application tasks people in specific places to collect specific pieces of data.

- | | | |
|--|-------------------------------------|--------------------------|
| 1. Beef | 12. Green Mustard | 25. Soybean |
| 2. Cassava | 13. Liquefied Petroleum Gas | 26. Spinach |
| 3. Chicken ("Village") | 14. Long Bean | 27. Sugar |
| 4. Chicken (Broiler) | 15. Mackerel | 28. Sweet Potato |
| 5. Chicken Eggs ("Village") | 16. Maize | 29. Tempeh |
| 6. Chicken Eggs (Broiler/Layer) | 17. Milk (Fresh) | 30. Tofu (Soft) |
| 7. Chillis (Green) | 18. Milk (Powdered) | 31. Tomato |
| 8. Chillis (Red) | 19. Noodles (Instant) | 32. Vegetable Oil |
| 9. Duck (Whole) | 20. Onion | |
| 10. Flour (Wheat) | 21. Peanuts (Raw, Unshelled) | |
| 11. Green Bean | 22. Rice (Low Quality) | |
| | 23. Rice (Medium Quality) | |
| | 24. Rice (Premium Quality) | |

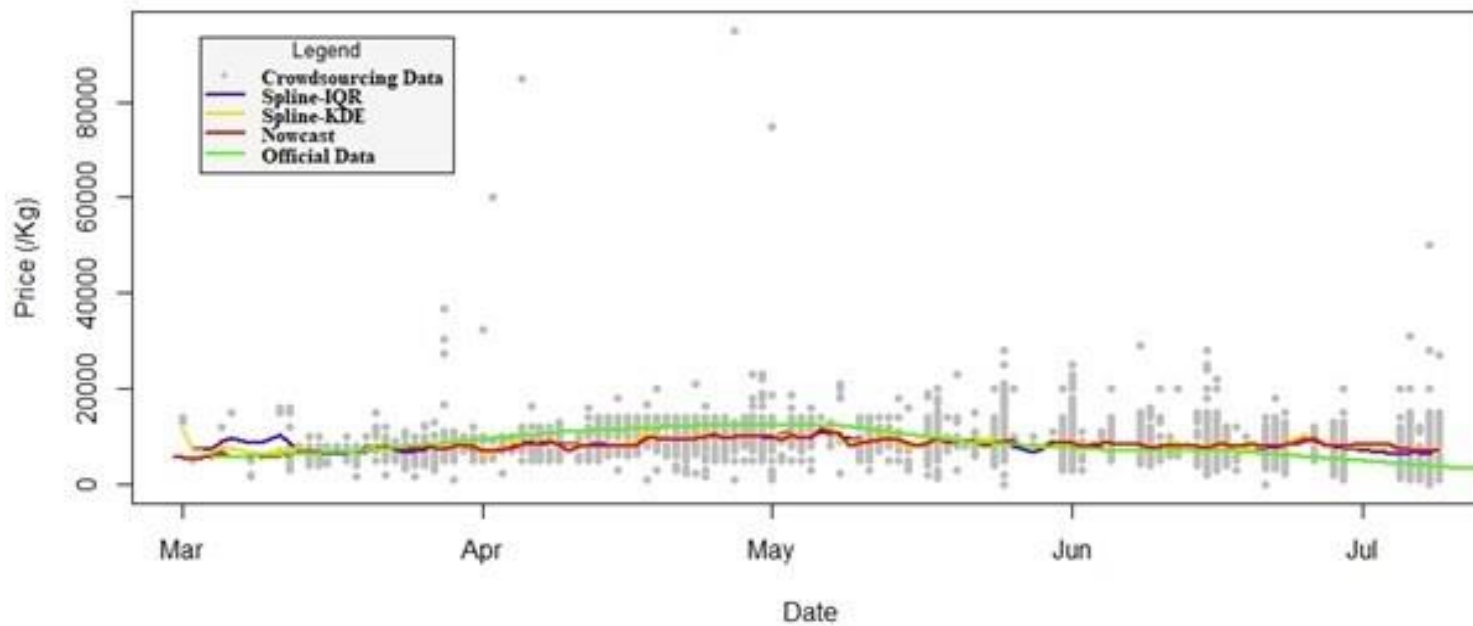
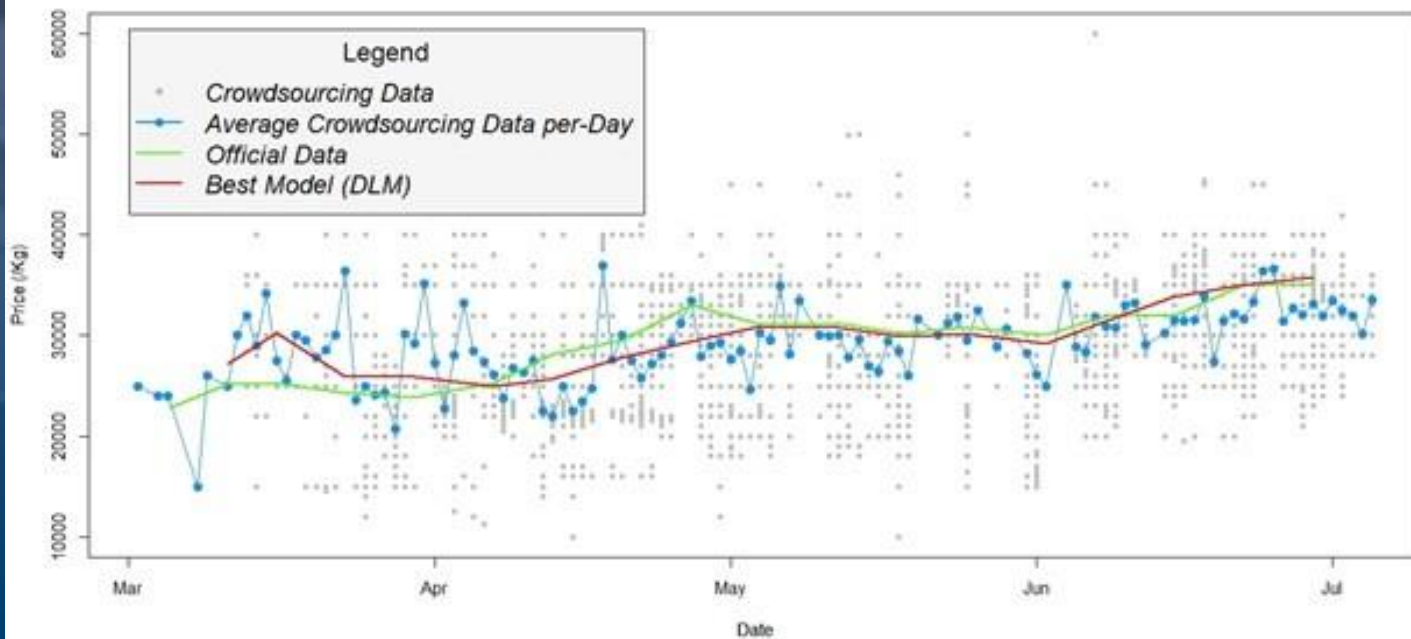
Table 2: The original list of 32 staples and the final list 20 staples (in bold).

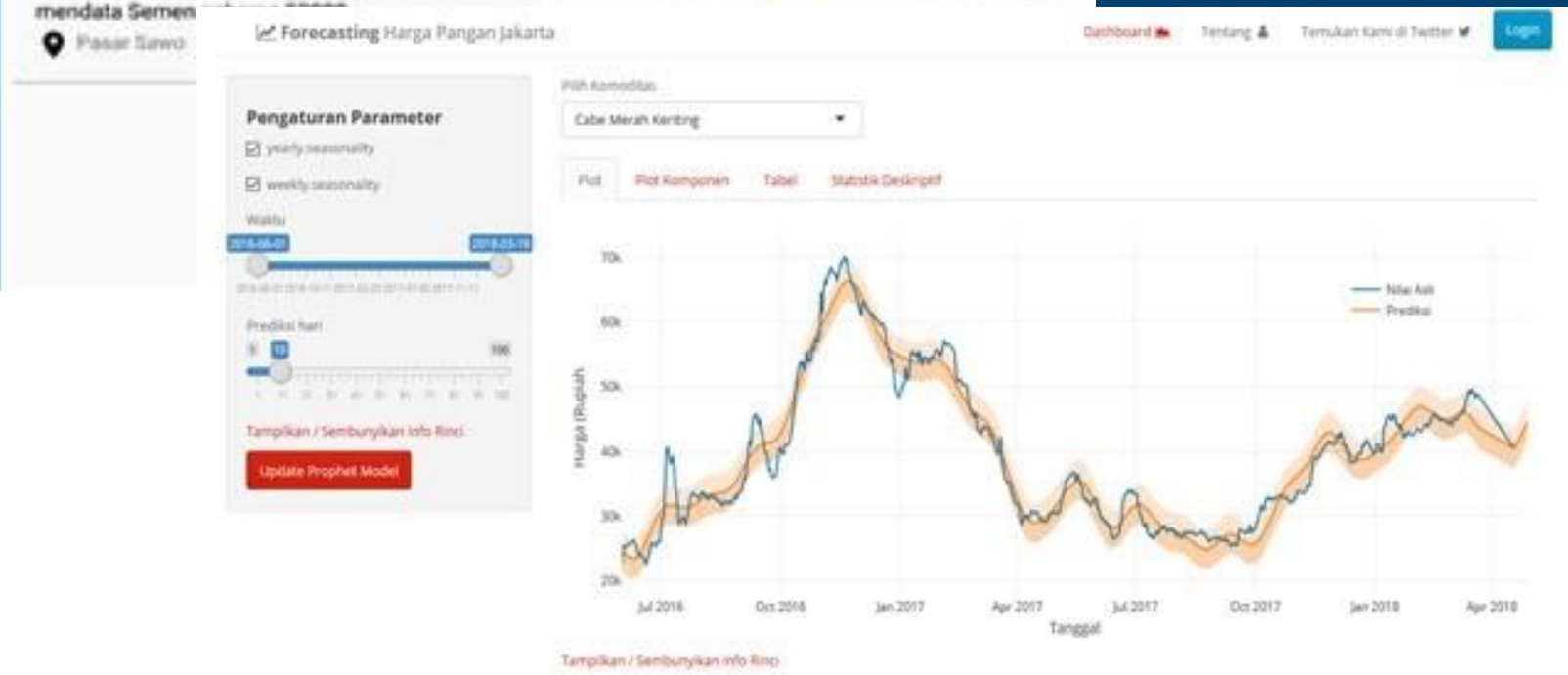


Number of daily reports of commodity fish in all markets period April-June 2015 in Lombok

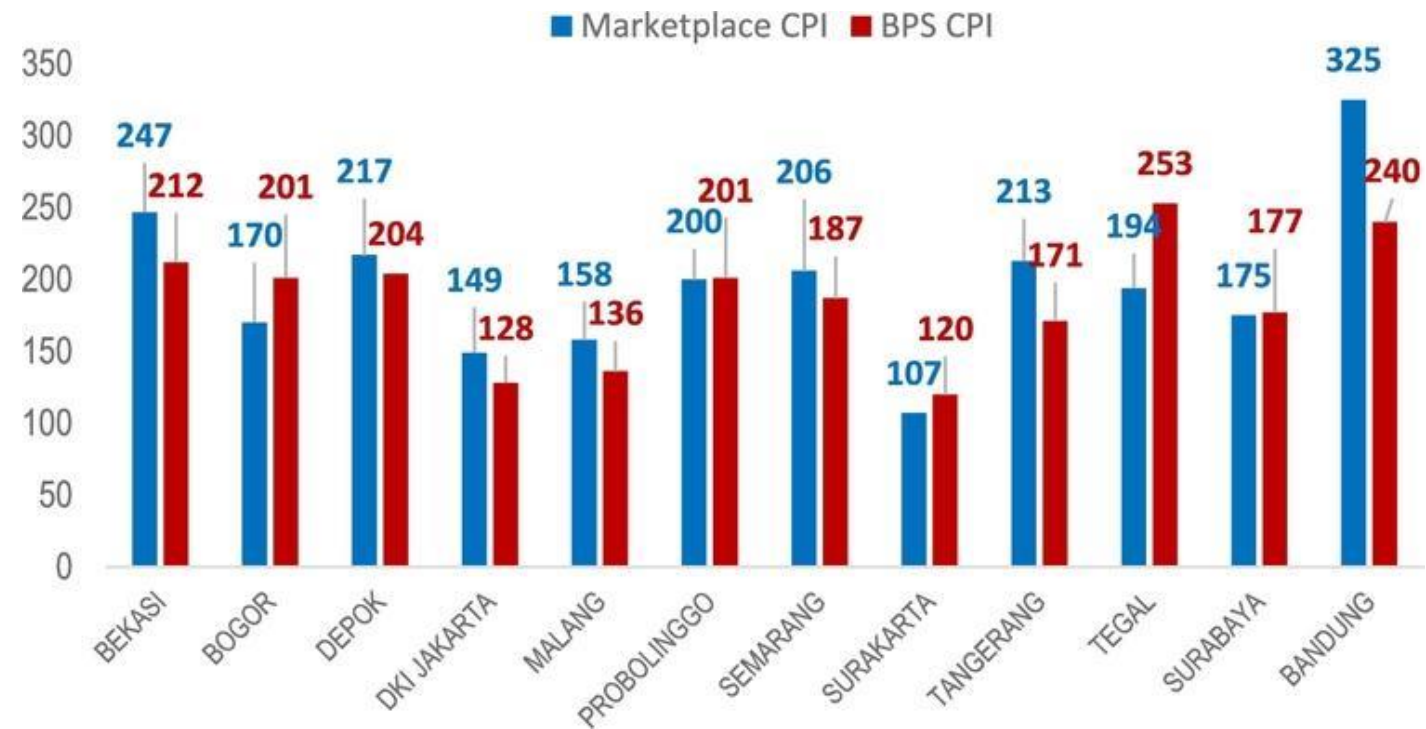


The Best Modeling Plot for Broiler Chicken Prices





CPB BASED ON MARKETPLACE



- *Use e-commerce data in calculating the CPI at city level.*
- *a case study of one of the largest online marketplaces in Indonesia*

Data Science 5 (2022) 79–95
DOI 10.3233/DS-210037
IOS Press

Implementation of marketplace data in the production of Consumer Price Index in Indonesia

Muhammad Khozy Al Haqqoni^a and Setia Pramana^{a,b,*}

^a *BPS-Statistics, Indonesia*

ORCID: <https://orcid.org/0000-0001-9239-1544>

^b *Politeknik Statistika STIS, Jakarta, Indonesia*

ORCID: <https://orcid.org/0000-0002-8590-1451>

Editor: Evangelos Pourmaras (<https://orcid.org/0000-0003-3900-2057>)

Solicited reviews: Cristina-Iulia Bucur (<https://orcid.org/0000-0002-7114-6459>); two anonymous reviewers

Received 31 March 2021

Accepted 1 October 2021

Big data implementation for price statistics in Indonesia: Past, current, and future developments

Setia Pramana^{a,b,*}, Siti Mariyah^a and Takdir^a

^a *Politeknik Statistika STIS Jakarta Indonesia*

^b *BPS Statistics Indonesia*

CPS2564 Tiffany Rizkika et al.



Nowcasting modelling of volatile and non-volatile food prices using crowdsourcing data (case study of some food commodities prices on Lombok island in 2015)



Tiffany Rizkika¹, Wida Siddhikara², Budi Yuniarto³, Ricky Yordani³, Setia

Pramana³

¹BPS Solok Selatan

²BPS Statistics Indonesia

³Politeknik Statistika STIS

Abstract

Monitoring of food prices in real-time or called nowcasting is important to maintain price transparency so that the government easy to detect inflation early and the economy becomes more stable and also can reduce the potential

- 
- <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200740>
 - <https://content.iospress.com/articles/data-science/ds210037>
 - [https://2019.isiproceedings.org/Files/10.Contributed-Paper-Session\(CPS\)-Volume-4.pdf](https://2019.isiproceedings.org/Files/10.Contributed-Paper-Session(CPS)-Volume-4.pdf)



Initiatives Big Data for Prices Statistics: Current Progress

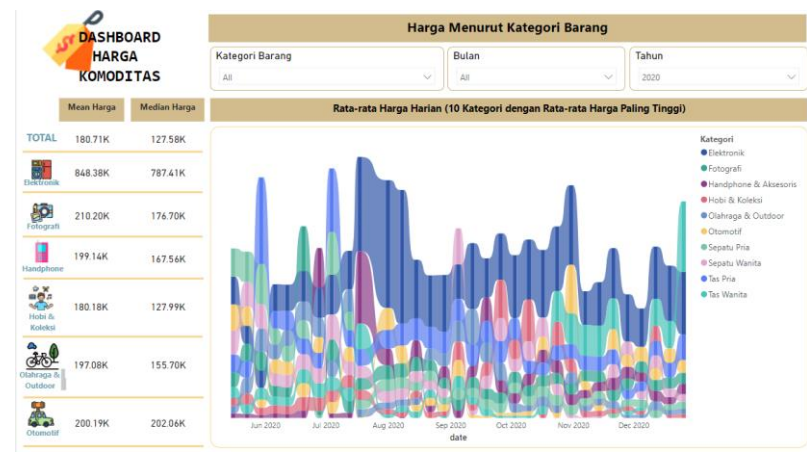
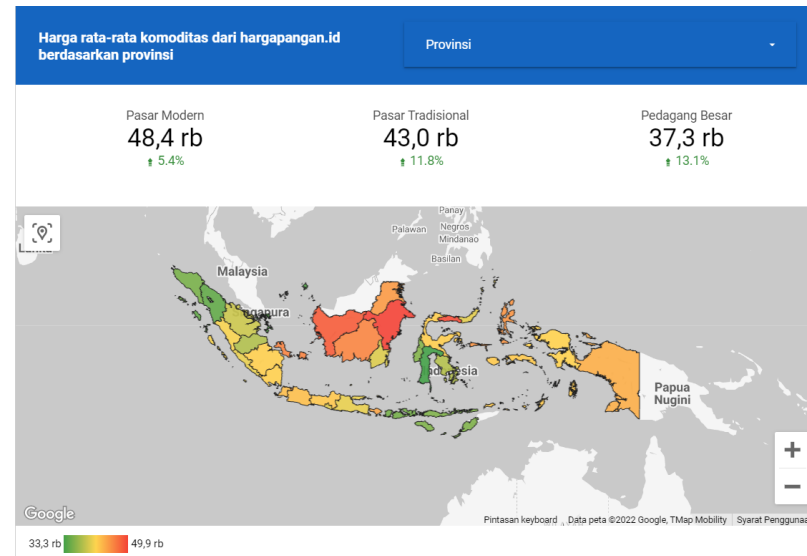
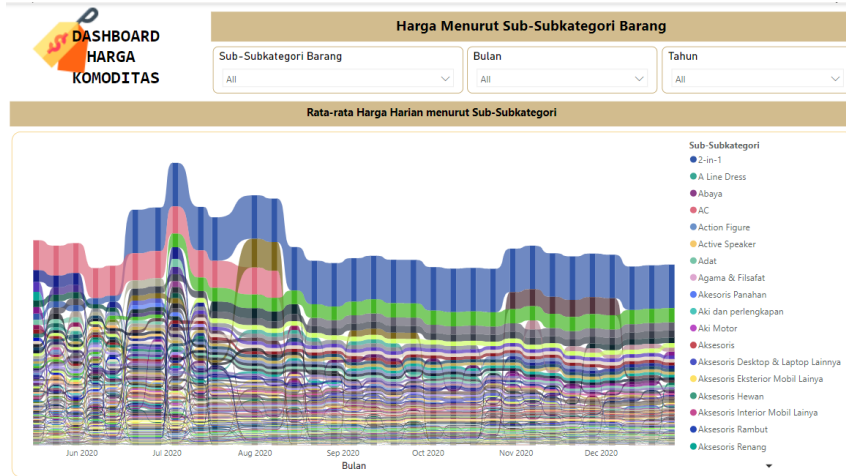
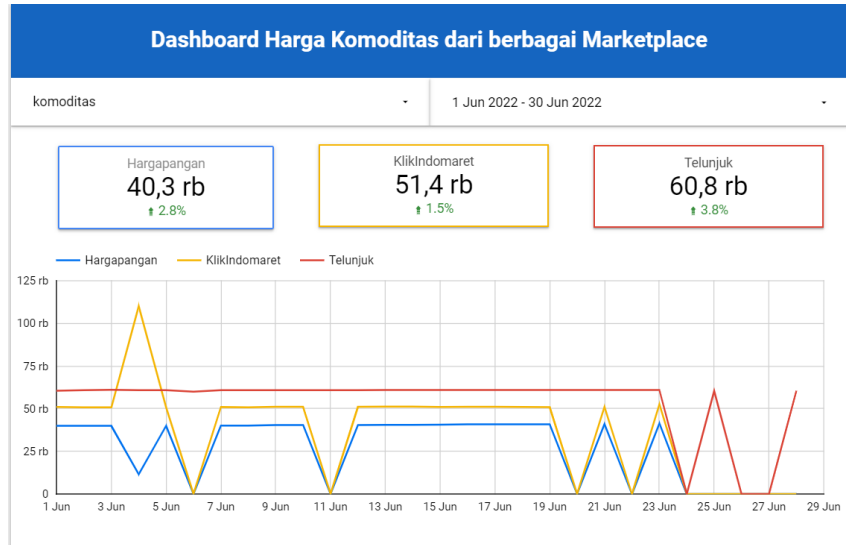
KOMODITAS HARGA DAN SUMBER BIG DATA



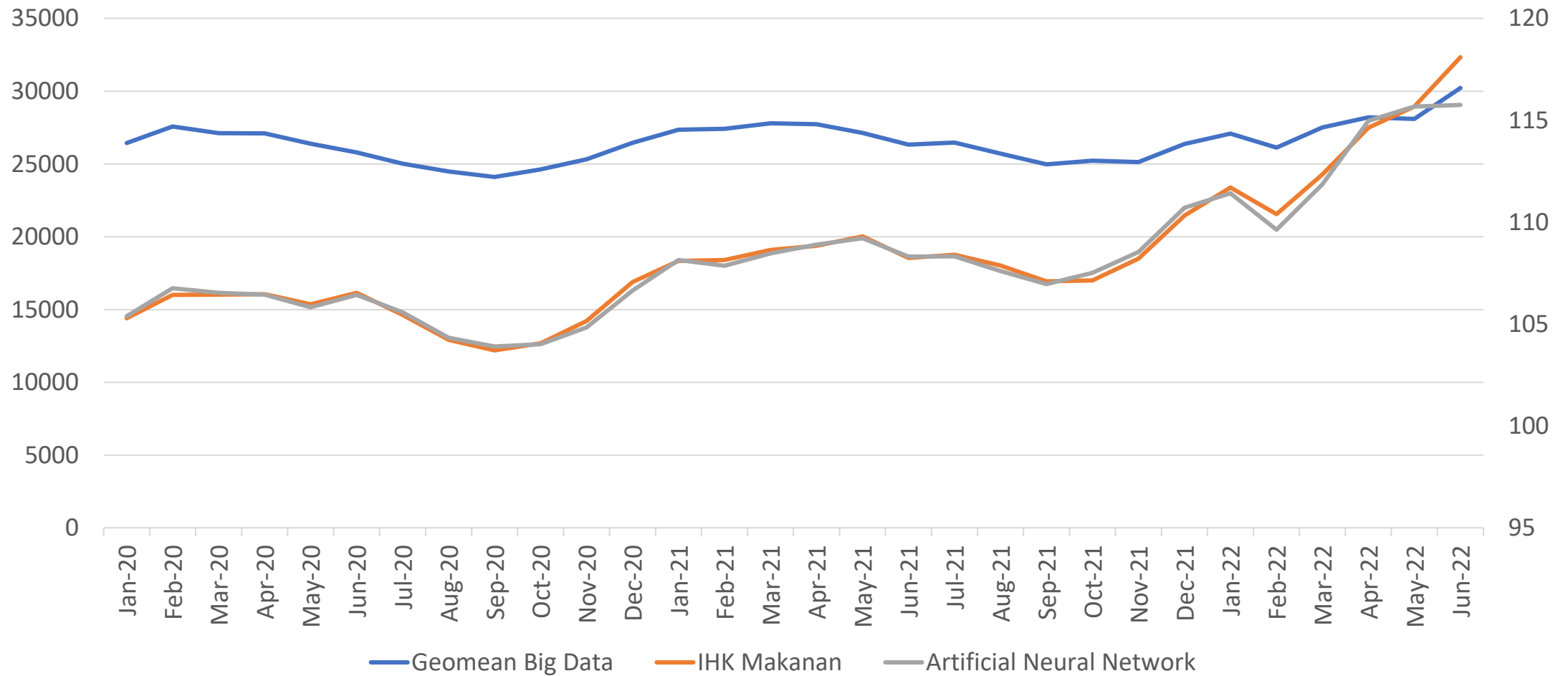
No.	Komoditas	Sumber Big Data	Bobot dalam IHK
1	Makanan, Minuman Dan Tembakau	Infopangan, Blibli, Shopee	26.01
2	Pakaian Dan Alas Kaki	Blibli, Shopee	5.35
3	Perumahan, Air, Listrik, Dan Bahan Bakar Rumah Tangga		19.68
4	Perlengkapan, Peralatan Dan Pemeliharaan Rutin Rumah Tangga	Blibli, Shopee	6.06
5	Kesehatan		2.66
6	Transportasi	Tiket.com	11.95
7	Informasi, Komunikasi, Dan Jasa Keuangan	Blibli, Shopee	5.34
8	Rekreasi, Olahraga, Dan Budaya	Blibli, Shopee	2.09
9	Pendidikan		5.65
10	Penyediaan Makanan Dan Minuman/Restoran		8.90
11	Perawatan Pribadi Dan Jasa Lainnya	Shopee	6.31

Data dari sumber Big Data kemungkinan dapat memproksikan 2 dari 4 pembobot besar dalam IHK

VISUALISASI DATA HARGA



PERMODELAN DATA HARGA



TANTANGAN PENGOLAHAN BIG DATA HARGA



Mendapatkan Satuan Unit

Dalam rencana membangun diagram timbang komoditas online, tentunya membutuhkan konversi yang benar dari komoditas. Namun, tidak semua komoditas di Shopee dan Blibli mencantumkan dengan jelas satuan unit yang sesuai



Pre-processing

Selain, menentukan satuan unit, pre-processing data harga memerlukan waktu yang cenderung lama dan memiliki variasi kemungkinan yang banyak dalam cleaning data



Storage

Kebutuhan akan penyimpanan sangat urgent pada big data harga karena cakupan dari komoditas online sangat banyak sehingga akan menggerus penyimpanan secara cepat



Terima Kasih!