Implementation of Marketplace Data in the Production of Consumer Price Index in Indonesia

Muhammad Ghozy Al HAQQONI a,[[1]](#footnote-1) and Setia PRAMANAa,b,[[2]](#footnote-2)

a BPS-Statistics Indonesia

b Politeknik Statistika STIS, Jakarta, Indonesia

**Abstract.** Digital Economy in recent years, especially in Southeast Asia, including Indonesia, is growing rapidly. E-commerce is one part of the Digital Economy. *Badan Pusat Statistik* (BPS-Statistics) as a Non-ministerial Government Agency responsible directly to the president has conducted an E-commerce Survey in 2019. From this publication it is concluded that the interest of Indonesian traders using the internet in selling in recent years has increased. So, the urgency of using e-commerce data in its application in official Statistics is increasingly needed. Several studies have carried out the application of e-commerce data in the calculation of Consumer Price Index (CPI). In this research, e-commerce data is applied with a case study using data from one of the online marketplaces in Indonesia in calculating CPI at the city level CPI in Java. The purpose of this study is to compare the marketplace CPI data and BPS-Statistics CPI. Data collection is done through web scraping techniques and followed by preprocessing data and analyzed descriptively. Web scraper that is built can be used in obtaining data. Commodity-level CPI with marketplace data tends to have relatively large prices which results in higher CPI being compared to BPS-Statistics CPI. While at the expenditure group level, the CPI between the two approaches is broadly similar in general.

**Keywords.** Consumer Price Index, Marketplace, Web Scraping

# Introduction

The digital economy is defined as a part of economic output that solely originates or primarily from digital technology with a business model based on digital goods or services [1]. In Southeast Asia, including Indonesia, the digital economy has developed rapidly in recent years. Anandan et al. predicted an internet economy of 200 billion USD in Southeast Asia in 2025 [2]. Furthermore, Anandan et al. saw such a drastic market acceleration that it predicted the internet economy to be closer to 240 billion USD [3]. This means that the role of the digital economy as a business transaction on the internet is now starting to be evenly distributed and will continue to grow.

Bank Indonesia (BI) stated that the emergence of the concept of the digital economy made economic actors have to adjust to changes that occur [4]. The Indonesian government has created an e-commerce roadmap [5] to realize Indonesia as the largest digital economy country in Southeast Asia by 2020. The roadmap in question is contained in Presidential Regulation Number: 74 of 2017 concerning the Road Map for the Electronic-Based National Trade System (Road Map E-commerce) 2017-2019. Quoting the contents of the Presidential Decree, the Presidential Regulation was issued with the consideration that an electronic-based economy has high economic potential for Indonesia, and is one of the backbones of the national economy, as well as in the context of optimizing the use of electronic-based economic potential. The government sees the need to encourage the acceleration and development of an electronic-based national trade system (e-commerce), start-up businesses, business development, and logistics acceleration by establishing an integrated E-commerce Road Map. .

The BEA (U.S. Bureau of Economic Analysis) classifies the digital economy into three main types of goods and services, namely infrastructure where digitization is needed so that it can be connected, e-commerce transactions, and digital media [6]. Mesenbourg [7] also divides the digital economy into three main components, namely supporting infrastructure, electronic business processes (e-business), and electronic commerce transactions (e-commerce). According to him, e-commerce is the value of goods and services sold on computers. E-commerce is defined as an online store and transaction system [8]. According to the Indonesian E-commerce Association (IdEA) in Wijaya & Mariyah's research [9], e-commerce is divided into three categories, namely (1) Classified ads, e-commerce where there is no online transaction facility from the website. (2) Marketplace, e-commerce where the website facilitates online transactions for online merchants. (3) Online retail, e-commerce where the website domain belongs to the online store itself and there is a stock of goods / services to be sold. An article published by Forbes [10] states that an online marketplace is a website or application that facilitates shopping from various sources. Marketplace operators don't own any inventory, their business is to serve other people's inventory to users and facilitate transactions.

BPS-Statistics Indonesia as a non-ministerial government agency that is directly responsible to the president has conducted the 2019 E-commerce Survey [11]. The survey was conducted on a sample of 3,504 Census Blocks spread across 101 cities / regencies in Indonesia. The survey results show that only 15.08 percent of respondents do e-commerce businesses. However, of the 15.08 percent of e-commerce businesses, 45.03 percent of businesses started using the internet from 2017 to 2018 and in 2019 as many as 25.11 percent, the rest started from 2016 onwards. From this it can be seen that the interest of Indonesian traders by using the internet in selling in recent years has increased. So that the urgency of using e-commerce data in its application to official Statistics is increasingly needed.

The study aim is to implement e-commerce data with a case study using data from one of the online marketplaces in Indonesia in calculating the CPI at the city level. A web scraper to obtain data from one of the online marketplaces in Indonesia was developed. From the data obtained, it is possible to compare the consumption value with the marketplace data sales value approach and the BPS-Statistics consumption value. Furthermore, the CPI obtained through the marketplace data approach is carried out with the BPS-Statistics CPI.

# Related Work

Several previous studies related to this research are used as references in research. The first research conducted by Wijaya & Mariyah [9], this study aims to capture price changes on e-commerce websites to produce e-commerce-based CPI through the development of a web scraper and evaluate the results of the e-commerce CPI with the CPI results from BPS-Statistics for each commodity. However, the research locus is assumed to be in the DKI Jakarta area due to limited online retail data which does not have information on sales locations. The results of this study indicate that the e-commerce-based CPI calculation can be done using the BPS-Statistics calculation approach

Likewise, Mithchell [12] uses collected online prices to generate CPI, Mithchell analyzes how online prices on consumer goods move relative to prices collected in retail stores in Canada and evaluates how e-commerce affects the CPI. and explore how the CPI can further adapt and develop in an ever-increasing digital world. The data source used is the price from the online website, namely Amazon Canada. According to Mithchell, price movements and the frequency of price changes at online prices are higher than outlets that have a physical presence as indicated by the relative price magnitude. Mithchell added that there is currently no consistent long impact on CPI at lower levels than the national level. Even so, the online price used for calculating the CPI can be used as an indicator in calculating the CPI because price recording can be done more quickly.

Additionally, Cavallo & Rigobon discusses the use of online prices on official Statistics, especially in calculating CPI [13]. The methodology used is in accordance with the concepts and formulas of calculating the official US Bureau of Labor Statistics CPI. The data source used comes from online retails collected from 25 countries. This data covers at least 70% of the commodity basket official Statistics. The data was collected using web scraping software which is built in Python and PHP programming languages. The software is in the form of a robot that can extract information from websites that have a similar structure. It is stated that the method used is data extraction through the HTML websites structure. This study explains the advantages and disadvantages of collecting data using web scraping techniques. The biggest advantage obtained with web scraping techniques is that it costs little, data is available in real time, and daily data is possible to obtain. Nonetheless, the disadvantage of data collection on online retail websites is that information is not available on many products sold. So that in calculating the CPI, adjustments are needed with those made by official Statistics.

# Consumer Price Index by BPS-Statistics

According to BPS-Statistics [14], the Consumer Price Index (CPI) is one of the economic indicators used to measure the rate of price change (inflation / deflation) at the consumer level. With changes in public consumption patterns, starting January 2020, the measurement of inflation in Indonesia will use the 2018 CPI base year . Several fundamental changes in the calculation of the CPI () compared to (CPI ), especially in terms of coverage, classification of commodity groupings, methodology for calculating the CPI, commodity baskets, and weigh diagrams.

These changes are based on the Cost of Living Survey (SBH) conducted by BPS-Statistics during 2018, as one of the main basic ingredients in calculating the CPI [14][15]. Where currently there are 835 commodity baskets used in each city covering between 248-473 types of commodity goods / services. The number of cities covered is 90 CPI cities. The classification of commodities at that time was divided into 11 groups and 34-42 sub-groups of consumption expenditures in the COICOP (Classification of Individual Consumption According to Purpose).

The CPI calculation by BPS-Statistics [14] in general is using the Modified Laspeyres method with the following formula:

|  |  |  |
| --- | --- | --- |
|  | , | (1) |

where:

= CPI at -th period,

= Price of -th type of goods/services at the -th period,

= Price of -th type of goods/services at period -th period,

= Consumption value of -th type of goods/services at the base year,

= Consumption value of -th type of goods/services at -th period,

= The number of types of goods / services listed in the commodity basket.

Furthermore, the BPS-Statistics [14] consumption value is obtained from the multiplication of commodity prices by the quantity (amount) consumed or can be formulated as follows:

|  |  |  |
| --- | --- | --- |
|  | , | (2) |

where:

= -th period of consumption value at -th commodity,

= Relative price of -th period at -th commodity,

= -th period of consumption value at -th commodity.

With relative prices [14] formulated as formula below:

|  |  |  |
| --- | --- | --- |
|  | , | (3) |

where:

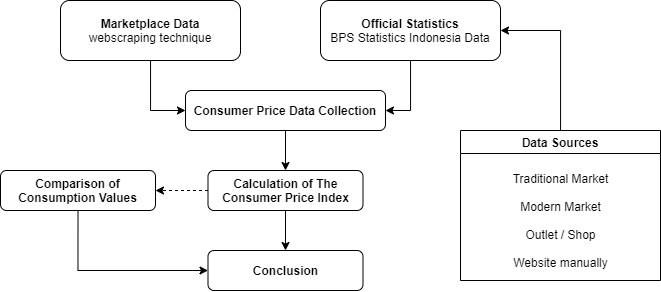
= Relative price of -th commodity at the -th month,

= geometric mean price per month of the -th period,

= geometric mean price per month for the -th period.

# Methods

This study examines the calculation of CPI with marketplace data using data collection methods in the form of web scraping techniques as an alternative to BPS-Statistics data collection methods. In addition, the consumption value sourced from BPS-Statistics is compared with the consumption value using the sales value approach obtained from marketplace data to see the consumption patterns of the people between the two approaches. The research framework in this study is visualized in Figure 1.



.**Figure 1.** Research framework.

## Data Collection

The data in this study were obtained through two approaches, the web scraping technique and the data obtained from the BPS-Statistics Subdirectorate of Consumer Prices by interviews and documents. The full web-scraped datasets remains available at <https://github.com/mghozyah/dataios>.

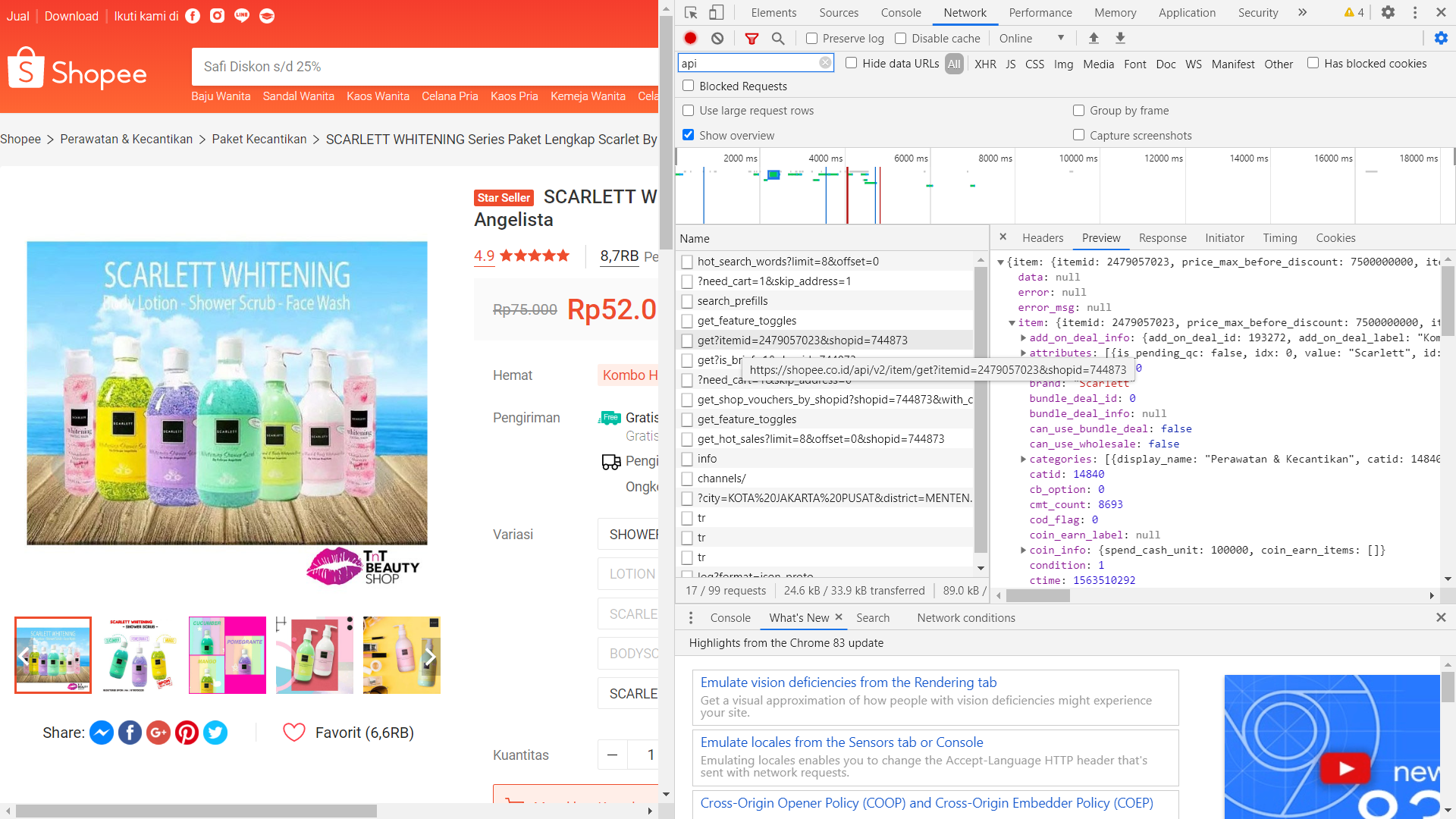
**Web Scraping**

According to Patel & Patel [16], web scraping is the process of collecting information from the web using a computer software program. A simple web scraper application scrapes web pages, extracts the required data from the response and then saves the extracted data to a data store such as a flat file. A web scraper usually consists of two main components with other components, namely:

* Web Request / Response Manager
  + Fetch data URL or credentials from data storage.
  + Make a request (request) to the targeted data source.
  + Take the response from the data source.
  + Directing the results of the response to the second component, namely the response parser.
* Response Parser
  + Fetches XPATH which contains the command to extract response data.
  + Extract the response data in XPATH and store the data on a medium such as a flat file.

Web scraping in the CPI Manual Concepts and Methods book published by the IMF [17] is defined as an automatic data collection method taken from the internet by extracting data from websites or using an API (Application Programming Interface), which is a series of routines, protocols, and tools for building software applications. Web scraping identifies and retrieves relevant data and organizes it in a format that suits your needs.

This study focuses on data collection based on selected commodities according to the BPS-Statistics commodity basket. So it is necessary to learn flow network or the infrastructure down to the product level. Figure 2 shows the flow network that will be used in scraping the marketplace data.

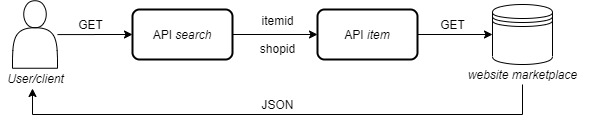


JSON Formatted Data

API *item*

**Figure 2.** The Flow Network of Researched Marketplace.

Figure 2 shows that the marketplace has well-organized website structure that the data is stored in JSON format and can be retrieved using public API, one of the API that shown in the Figure 2 named the API item. The data collection flow with web scraping techniques is shown in Figure 3.

 **Figure 3.** Flow of request data in web scraper

**Interviews**

An interview to the following respondent was carried out:

* Staff of the Sub-directorate for Statistical Model Development to discuss the potential use of marketplace data at BPS-Statistics and to understand the calculation of CPI carried out by BPS-Statistics.
* Staff of the Sub-Directorate of Consumer Price Statistics to deepen the stages of calculating the BPS-Statistics CPI, obtain information on the process of handling data cases, obtain some documents needed in the study, and obtain information about the Consumer Price Index approach using online enumeration data that has been conducted by BPS-Statistics
* *Politeknik Statistika STIS* staff who are members of *Politeknik Statistika STIS* Big Data team to learn how to collect research marketplace data which is currently also being carried out by the team.

**Documents**

The documents used are documents originating from the Sub-directorate of Consumer Price Statistics and published by BPS-Statistics. There are seven documents used as a reference, to be precise:

1. The 2013 Consumer Price Survey (SHK) Guidebook [14].

This handbook is used to study the concepts and definitions and methodology of consumer price surveys in collecting consumer price data.

1. 2019 BPS-Statistics E-commerce Statistics [11].

The results of this publication are used to learn the development of e-commerce business conditions in Indonesia.

1. Book 2020 Consumer Price Index Processing, Book 5 [14].

This book is used to review the method of calculating the CPI used by BPS-Statistics.

1. List of Commodity Baskets according to the 2018 SBH [15].

The commodity basket list is used as a list of keywords in retrieving data using web scraping techniques. This list is also used as a means of matching commodity data obtained with commodity baskets available in certain CPI cities. The results of matching these commodities will be used in calculating the CPI for each CPI city on Java Island according to the 2018 SBH. The commodity list contains information on the code of the expenditure group, expenditure group, expenditure sub-group code, expenditure sub-group, commodity code, commodity, city and city code.

1. CPI Weigh Diagram for 2018 SBH Results, Book 2 [18].

This document is used as a reference for basic consumption values ​​with the new base year (2018 = 100) used in calculating the research CPI.

1. List of Consumption Value (CV) in Java Island in April and May 2020.

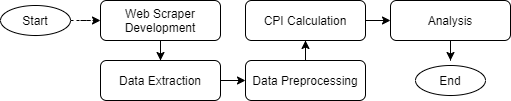
This list of values ​​is used as the fulfillment of the criteria for calculating the CPI for commodities using the Modified Laspeyers method because the CPI Weighing Chart for SBH Results in 2018, Book 2 does not contain the Q0 value separately and the consumption value approach with marketplace data is different from the BPS-Statistics concept and definition.

1. May 2020 CPI for commodities in each CPI city on the island of Java.

This list is used to compare the results obtained and the official statistics produced by BPS-Statistics Indonesia with approaches to different types of markets, namely online and offline.

## Data Analysis

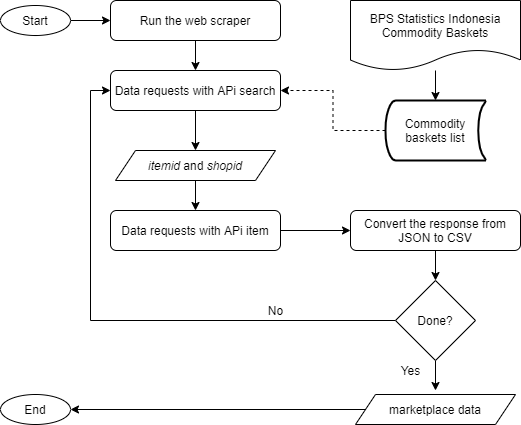
The analytical method used in this research is descriptive analysis. The analysis was carried out after the data were ready to be analyzed. The flowchart of the research stages carried out in this study (Figure 4).



**Figure 4.** Research Analysis Flow

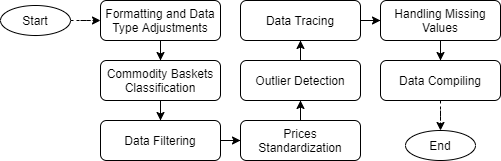
**Data Extraction**

The web scraper that has been built will be run by searching for commodity data based on BPS-Statistics commodity baskets, whereas a list of commodity basket to be included in the search is required. The BPS-Statistics Indonesia commodity basket according to the 2018 SBH (Cost of living Survey) consists of 11 Expenditure Groups which consist of commodity goods and services. The data that has been obtained need to be extracted into flat file for analysis purposes. The flow of how to extract data in this study can be seen in Figure 5. The extracted data through scraping then entered in the data preprocessing stage as follow in Figure 6.



**Figure 5.** Data Extraction Flow

**Data Preprocessing**



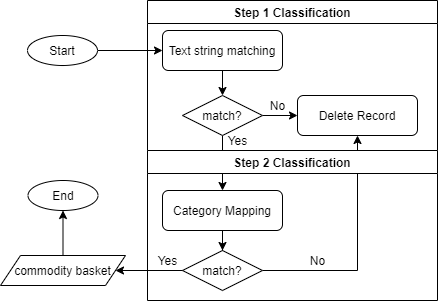
**Figure 6.** Data Preprocessing Flow

* Formatting and Data Type Adjustments

Data adjustments were made due to errors in the data format when the data was collected.

* Commodity Basket Classification

This research classification technique refers to the web scraping data classification technique in the CPI Manual Concepts and Method [17]. There are two techniques for classifying commodities in research. First text string matching and second category mapping. Two-technique classification is carried out to avoid misclassification of commodities because product names are variables that are inputted manually by the user, while category variables are inputted automatically by the system. The following figure shows how the classification works in this study.



**Figure 7.** Commodity Basket Classification Flow

* Data Filtering

BPS-Statistics in selecting respondents chose merchants with the criteria of being crowded with consumers. In the marketplace data approach, data that has no history / trace of the seller means that it cannot be used as an object of analysis. So at this stage the data records that have no sales footprint are filtered, in this study the data records are filtered using the *"hist\_sold"* variable.

Furthermore, the data was filtered based on location. First, eliminating data records that are located outside Indonesia and data records without location information. Next, filtered the location of the data records that entered the CPI city in Java, where new variables had previously been formed to support the filtering stage, namely the *"information"* variable (Not CPI City on Java Island; CPI City on Java Island) and *"cpi\_city".* (Name of City / Regency in Java Island).

* Price Standardization

Standardization of the price of a record is carried out to equalize the size of a unit [19]. According to the concept and definition of BPS-Statistics itself, retail traders use transactions with standard units such as kilograms and liters. So that in this study, the price standardization of each record was carried out into a standard unit by performing one of the web-scraped data classification techniques [17], namely manual mapping.

* Outlier Detection

Complete a selection of records detected as outlier data to avoid fraudulent data and the effect on the standardized average price [17][20]. This method is used based on IMF guidelines in dealing with outliers data. It is stated that prevention is seen based on the median and the quartile value. Thus, in this study the control of outlier data used the IQR method [21].

* Data Tracing

Data tracing is done to filter data that has price movements in each month. At this stage, the “*weeks*” and “*month*” variables are formed which are used as additional information on the time span of a data record and preparation for calculating the Consumer Price Index.

* Handling Missing Values

The elimination of extreme prices is also carried out to avoid the impact that occurs on average commodity prices [20]. In the CPI Manual Concepts and Method (2020) book, it is stated that there are four ways to handle lost data [17], that is:

1. Ignore observations for which price data are missing so that a suitable sample is retained even though there are fewer samples.
2. Uses the last observed price on the missing price data.
3. Imputation of missing price data with changes in average prices over a certain period.
4. Imputation of missing price data based on price changes that are comparable to other similar outlets.

Based on the results of interviews with informants, currently the handling of lost data at BPS-Statistics uses a one-month waiting system, where if there is some missing data, the price is equated with the last census. In the second month, if nothing is found, a change is made by asking for the price in the previous month. So, in this study, the prevention of missing data was carried out using the Last Observation Carried Forward (LOCF) method.

LOCF is one of the missing data imputation methods that are usually used in longitudinal data studies [22]. LOCF was used to maintain sample size as well as to reduce bias caused by attrition of participants in the study.

* Data Compiling

The data from the preprocessing of the previous stages, compiled the research data with the data obtained from the Sub-directorate of Consumer Price Statistics by taking slices of commodity baskets. So that it will simplify the process of calculating the CPI and analysis.

**CPI Calculation**

The data that has been through the data preprocessing stage is then calculated to the CPI. The method used is to use a modified Laspeyers index [14]. In this study, several changes were made from the BPS-Statistics processing stage. The difference between the marketplace and BPS-Statistics processing can be seen in the following table.

**Table 1.** Differences in BPS-Statistics CPI Processing and Marketplace CPI

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Differentiator** | **BPS-Statistics** | **Marketplace** |
| 1 | Market Weight | Location consists of the name of the market within a market type. However, current weighting has only been applied to the rice commodity | Market location listing was not carried out because the selection of respondents on marketplace data was different from the selection of BPS-Statistics respondents.  Therefore, the weight of the location is neglected. |
| 2 | Weighted by Type of Market | Based on the weight of traditional markets and modern markets. | Because it only consists of one type of market, namely the online marketplace , the weight of location type is 1 which in other words can be ignored.. |
| 3 | Quality Weight | Calculating the average commodity price for each quality. | Calculation of the average commodity price is by calculating the average price of all qualities with similar commodities. |
| 4 | Commodity Weights | Used in processing CPI at the national level.. | This study it did not cover all CPI cities according to the 2018 SBH. So, the calculation of the CPI is limited to the city level of the CPI, |
| 5 | Consumption Value | The Basic Consumption Value uses the BPS-Statistics Base NK for the 2018 SBH results. | The value of total marketplace and BPS-Statistics consumption by expenditure group is adjusted to the number of commodities covered in the research. This consumption value is used in calculating the CPI for the expenditure group. |

**Analysis**

The results of calculating the CPI in the previous stage are used for analysis at this stage. The calculation of the average price is carried out in April and May 2020, so that the results of the calculation are in the form of May 2020 CPI for commodity and per group with adjustments. At this stage, the descriptive analysis method is used. Descriptive analysis is presented with tables and graphs to evaluate the research results.

# Results and Discussion

## Web Scraping Implementation

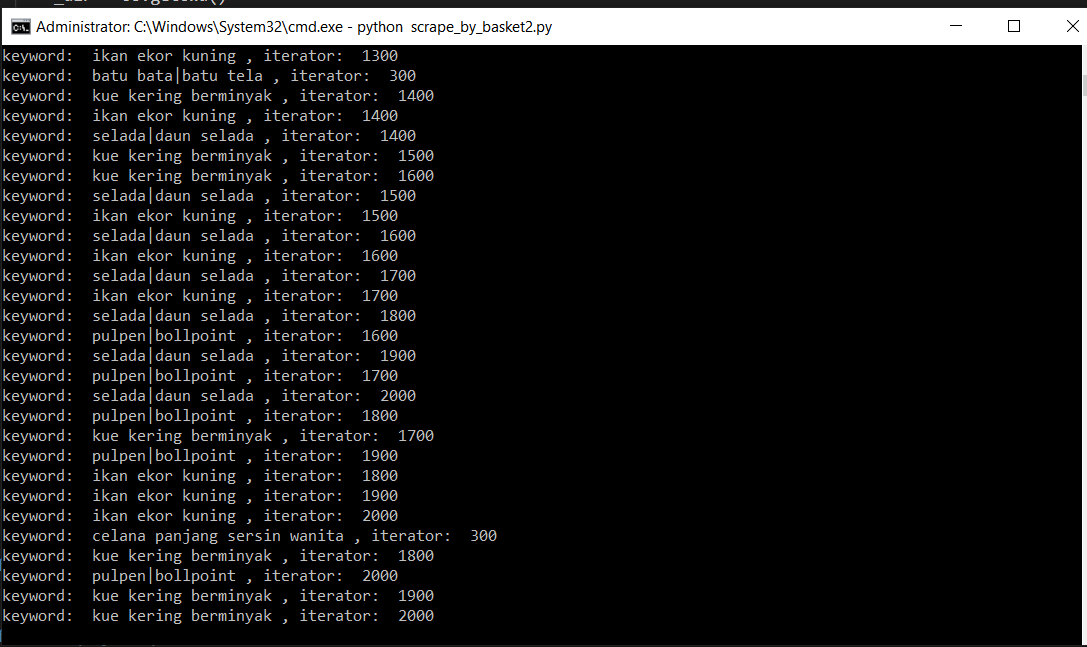
**Listing Commodities:** This study uses a list of commodity items included in the BPS-Statistics commodity basket as a result of the 2018 Cost of Living Survey (SBH). The BPS-Statistics commodity basket consists of 835 types of commodities from 11 expenditure groups, where the expenditure groups with codes 09 and 10 are service commodities, which is not the scope of this study. The details of the expenditure groups covered in the study are presented in following table.

**Table 2.** List of Expenditure Groups in this Study

|  |  |  |
| --- | --- | --- |
| **No** | **Group Code** | **Expenditure Group** |
| 1 | 01 | Food, Beverages and Tobacco |
| 2 | 02 | Clothing and Footwear |
| 3 | 03 | Housing, Water, Electricity, and Household Fuels |
| 4 | 04 | Equipment, equipment and routine Household maintenance |
| 5 | 05 | Health |
| 6 | 06 | Transportation |
| 7 | 07 | Information, Communication and Financial Services |
| 8 | 08 | Recreation, Sports and Culture |
| 9 | 11 | Personal Care and Other Services |

The results of commodity listings used for search on the web scraper amounted to 535 types of commodities.

**Running The Web Scraper:** The list of listing results in the previous stage is used as a marketplace data search suggestion. The web scraper that has been built is made by scheduling that is ready to run automatically. In this study, data collection was scheduled for each day. The following is the interface when the web scraper is running shown in Figure 8.



**Figure 8.** Data retrieval process with this research web scraper

At each iteration, the JSON response that is obtained, is extracted into CSV format. If there is a problem in the scraping process such as a lost connection, the data that has been obtained is still stored.

## Research Commodity Coverage

The data that has been compiled are then tabulated to see the coverage of the final commodity in the study. The final commodity coverage of this research can be seen in the following table.

**Table 3.** BPS-Statistics and this study commodities coverage finale

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **CPI City** | **Province** | **BPS-Statistics** | **Marketplace Coverages** | **Final Coverage** |
| 1 | DKI Jakarta | DKI Jakarta | 473 | 470 | 221 |
| 2 | Bogor | West Java | 351 | 262 | 97 |
| 3 | Sukabumi | West Java | 333 | 82 | 18 |
| 4 | Bandung | West Java | 411 | 367 | 187 |
| 5 | Cirebon | West Java | 374 | 165 | 55 |
| 6 | Bekasi | West Java | 350 | 364 | 127 |
| 7 | Depok | West Java | 341 | 371 | 121 |
| 8 | Tasikmalaya | West Java | 355 | 152 | 48 |
| 9 | Cilacap | Central Java | 306 | 114 | 20 |
| 10 | Purwokerto | Central Java | 364 | 0 | 0 |
| 11 | Kudus | Central Java | 314 | 134 | 13 |
| 12 | Surakarta | Central Java | 356 | 285 | 99 |
| 13 | Semarang | Central Java | 408 | 295 | 117 |
| 14 | Tegal | Central Java | 340 | 81 | 28 |
| 15 | Yogyakarta | DIY Yogyakarta | 400 | 227 | 93 |
| 16 | Jember | East Java | 358 | 157 | 34 |
| 17 | Banyuwangi | East Java | 315 | 109 | 22 |
| 18 | Sumenep | East Java | 320 | 34 | 4 |
| 19 | Kediri | East Java | 373 | 154 | 61 |
| 20 | Malang | East Java | 382 | 281 | 113 |
| 21 | Probolinggo | East Java | 340 | 67 | 22 |
| 22 | Madiun | East Java | 335 | 124 | 33 |
| 23 | Surabaya | East Java | 420 | 382 | 162 |
| 24 | Tangerang | Banten | 335 | 364 | 132 |
| 25 | Cilegon | Banten | 333 | 114 | 27 |
| 26 | Serang | Banten | 375 | 130 | 44 |

Table 3 shows the number and distribution of commodities covered in the study. These commodities are used in the calculation of the CPI to analysis. Purwokerto in this study cannot be analyzed because there is no commodity packages covered and DKI Jakarta has the highest final coverage, which is 221 commodities. This is in line with expectations because that DKI Jakarta is the capital of Indonesia.

## Consumption Value Patterns by BPS-Statistics and Marketplace Data

The marketplace consumption value is obtained from the processing results at the CPI calculation stage, while the BPS-Statistics consumption value is obtained from the Sub-directorate of Consumer Price Statistics. The consumption value used to see the movement pattern is at the level of the expenditure group.

In this section, a random sample is taken in the format of expenditure groups in a CPI city for each particular movement pattern (Figure 9) where there are 162 events with 9.26% or 15 events experiencing conditions where the BPS-Statistics Consumption Value (CV) is lower than the marketplace CV. In the following figure, the CV data was first carried out with a 10 logarithmic transformation to show the movement pattern more clearly because some incidents of difference were very small. Complete data on NK can be seen in Appendix section.

|  |  |
| --- | --- |
| **BPS-Statistics and Marketplace CV Patterns Increases**  **(28 events)** | **BPS-Statistics and Marketplace CV Patterns Decreases**  **(10 events)** |
|  |  |
| **BPS-Statistics CV Patterns Increases but Marketplace Patterns Decreases (23 events)** | **BPS-Statistics CV Patterns Decreases but Marketplace Patterns Increases (30 events)** |
|  |  |
| **BPS-Statistics CV Patterns has No Changes but Marketplace Patterns Increases (26 events)** | **BPS-Statistics CV Patterns has No Changes but Marketplace Patterns Decreases (45 events)** |
|  |  |
| description: : BPS-Statistics *Marketplace* | |

**Figure 9.** Conditions of Consumption Values (CV) by BPS-Statistics and Marketplace

From Figure 9, it can be seen that there are 38 events or 23.46% events of the consumption value that have a similar pattern, there are 53 events or 32.72% contradictory patterns, and the rest or 43.82% BPS-Statistics does not change when the marketplace changes. This means that the movement of marketplace data consumption patterns is more volatile than the pattern of consumption values obtained from BPS-Statistics.

## Marketplace Consumer Price Index Results with BPS-Statistics

Differences in the concepts and definitions by BPS-Statistics, such as the weight of quality and consumption value for calculating the CPI and the unavailability of marketplace supporting data are the main limitations in this study. Nonetheless, after going through the adjustment process and approach, this research was able to produce CPI with a marketplace data approach.

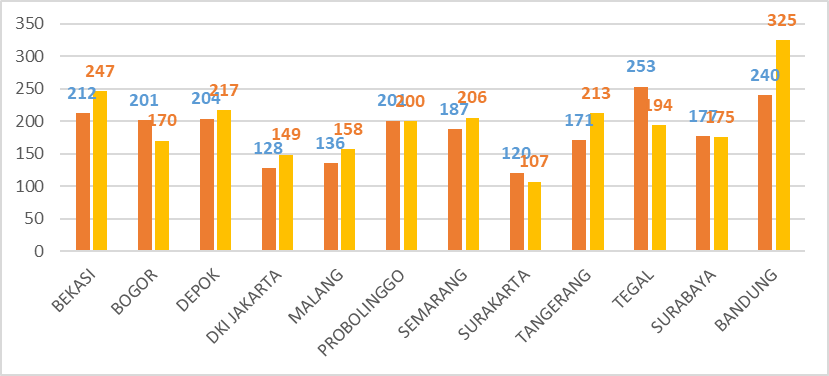
Furthermore, a descriptive analysis was carried out regarding the CPI that had been produced in the study. The analysis was carried out in each city and categorized into two groups, namely at the commodity level and the level of the expenditure group. The CPI that was compared in the study was the CPI for May 2020.

**Commodity Level**

At the commodity level, two cases were analyzed based on the BPS-Statistics CPI value, the commodities that generally has the highest and lowest BPS-Statistics CPI.

1. Highest BPS-Statistics CPI Value in General

Based on the May 2020 BPS-Statistics CPI, the commodity that generally has the highest CPI value in each CPI city is the commodity of shallots (Figure 10). Cities with shallot commodities as commodities with the highest CPI are Bandung, Bekasi, Bogor, Depok, Probolinggo, Semarang, Tangerang, and Tegal, where the data on the marketplace data for shallot commodities is covered by 12 CPI cities. The other four cities are DKI Jakarta with Shallot CPI in 14th place, Malang 3rd, Surabaya 2nd, and Surakarta 4th.



description: MarketplaceCPI BPS-Statistics CPI

**Figure 10.** BPS-Statistics and marketplace CPI for shallot commodities.

Figure 10 shows that the marketplace CPI at the commodity level generally has a higher index than the BPS-Statistics CPI. Five cities that had lower index scores than BPS-Statistics CPI were Bogor, Probolinggo, Surakarta, Tegal, and Surabaya. Based on several online news publications, the demand for shallots before Eid al-Fitr (Islamic festival) has experienced a rapid increase in prices. This incident caused a high CPI for both the BPS-Statistics CPI and the CPI with a marketplace data approach

1. Lowest BPS-Statistics CPI Value

The commodities of each city with the lowest BPS-Statistics CPI consisted of various commodities, the lowest was shown in the city of Malang with a value of 43.86 for the Red Chili commodity. The lowest BPS-Statistics CPI ranges from 43.86 to 103.21 CPI units (Table 4).

**Table 4.** Summary of BPS-Statistics and Marketplace CPI in each city with the lowest BPS-Statistics CPI as base in May 2020

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **CPI City** | **Commodity** | **BPS-Statistics CPI** | **Marketplace CPI** |
| 1 | Bandung | *Jengkol* | 66,47 | 77,32 |
| 2 | Banyuwangi | Bleach | 102,77 | 106,38 |
| 3 | Bekasi | Leek | 71,93 | 68,20 |
| 4 | Bogor | Deodorant | 75,85 | 77,21 |
| 5 | Cilacap | Women's shoes | 100,44 | 101.09 |
| 6 | Cilegon | Salt | 96,67 | 87,95 |
| 7 | Cirebon | Pepper / Pepper | 71,52 | 117,15 |
| 8 | Depok | Cayenne pepper | 76,18 | 119,64 |
| 9 | DKI Jakarta | Soun | 71,17 | 71,34 |
| 10 | Jember | Towel | 84,12 | 92,97 |
| 11 | Kediri | Avocado | 49,21 | 47,65 |
| 12 | Kudus | Panties | 101,76 | 99,70 |
| 13 | Madiun | Garan | 88,97 | 90,71 |
| 14 | Malang | Red chili pepper | 43,86 | 60,97 |
| 15 | Probolinggo | Garlic | 55,64 | 70,27 |
| 16 | Semarang | Cayenne pepper | 75,83 | 107,29 |
| 17 | Serang | Snack / Snack | 97,05 | 113,05 |
| 18 | Sukabumi | Towel | 77,44 | 85,10 |
| 19 | Sumenep | Potato | 103,21 | 102,47 |
| 20 | Surabaya | Avocado | 74,94 | 69,20 |
| 21 | Surakarta | Salt | 87,14 | 76,85 |
| 22 | Tangerang | Red chili pepper | 79,9 | 113,20 |
| 23 | Tasikmalaya | Cayenne pepper | 74,43 | 93,65 |
| 24 | Tegal | Salt | 89,93 | 85,63 |
| 25 | Yogyakarta | Cayenne pepper | 84,23 | 98,90 |

The BPS-Statistics CPI tends to be lower than the marketplace CPI. Cayenne pepper is a commodity that appears the most as the lowest commodity in several CPI cities, namely Depok, Semarang, Tasikmalaya, and Yogyakarta. Meanwhile, the CPI with the marketplace data approach in that city has a large enough gap. This means that price changes in marketplace data for cayenne pepper are greater than price changes with traditional and modern market approaches for calculating the BPS-Statistics CPI.

**Expenditure Group Level**

At this level, comparisons are made between expenditure groups in each city, so that differences in the amount of the CPI value can be analyzed using the BPS-Statistics and marketplace approaches. In this paper, two samples from 25 cities were taken with the highest number of commodity coverage Jakarta, and the lowest was Sumenep. The rest comparison results can be found in the appendix.

**DKI Jakarta:** DKI Jakarta's CPI at the expenditure group level tends to be the same. The figure below (Figure 11) shows that the difference in the CPI using the two approaches in each expenditure group does not exceed 15 CPI units. The biggest difference occurred in the Food, Beverage and Tobacco (1) expenditure group, which was 14.43 CPI units.

description: MarketplaceCPI BPS-Statistics CPI

**Figure 11.** DKI Jakarta’s BPS-Statistics and marketplace CPI at Expenditure Group Level.

**Sumenep:** The CPI of Sumenep City in the Food, Beverage and Tobacco expenditure group (1) if calculated using the BPS-Statistics data approach and marketplace data has a very small difference, that is 1.11 CPI units (Figure 12).

description: MarketplaceCPI BPS-Statistics CPI

**Figure 12.** Sumenep’s BPS-Statistics and marketplace CPI at Expenditure Group Level.

# Conclusion and Future Works

Based on the results and discussion that has been done, the following conclusions can be obtained:

1. The construction of a web scraper in one of the marketplaces in Indonesia to obtain marketplace data has been successfully built. The web scraper created can extract data directly into CSV format after receiving a response from the marketplace website. However, it is necessary to pay attention to internet connection when obtaining the data.
2. The consumption value using the marketplace data sales value approach tends to be lower than the BPS-Statistics consumption value. However, the patterns of consumption value movement in general are similar.
3. The CPI with a marketplace data approach at the city level tends to have a higher index value than the BPS-Statistics CPI. This indicates a difference in relative prices for BPS-Statistics commodities and marketplace data. At the expenditure group level, the BPS-Statistics CPI and the marketplace in general have quite a small difference, meaning that the CPI calculation with marketplace data is quite good.

These are some suggestions for future works to continue this research:

1. Commodities produced in this study do not include all commodity items listed in the 2018 SBH commodity basket. This is due to the classification of commodities using text string matching and category mapping. For this reason, in further research, it can be studied the classification method in marketplace data to obtain maximum results.
2. The treatment of standardization of prices on commodities also still uses manual mapping, so it is necessary to study further how to detect the quantity of a unit so that the process and results of standardization are more precise.

# Acknowledgements

This research would not have been possible without the support from the interviewees who have agreed to be interviewed for the purposes of this research, especially staff of the Sub-directorate for Statistical Model Development BPS-Statistics, staff of the Sub-Directorate of Consumer Price Statistics BPS-Statistics, and *Politeknik Statistika* STIS.

# Appendix

The full datasets of web-scraped data, marketplace consumption values, and the marketplace CPI’s are remains available at <https://github.com/mghozyah/dataios>

References

1. Bukth, R. & Heeks, R. (2017). Defining, Conceptualising and Measuring the Digital Economy. Manchester Centre for Development Informatics Working Paper 68. Manchester, UK. Available at <https://www.research.manchester.ac.uk/portal/en/publications/defining-conceptualising-and-measuring-the-digital-economy(c9ae8345-6f3c-4d5e-9496-d7bd00153df3)/export.html>
2. Anandan, R., Sipahimalani, R., Bharadwaj, Alap., Jhangiani J., Kim D., Ramesh, S. (2016). e-conomy SEA: Unlocking the $200B Digital Opportunity. Asia Pasific: Google-Temasek. Available at <https://www.thinkwithgoogle.com/intl/en-apac/future-of-marketing/digital-transformation/e-conomy-sea-unlocking-200b-digital-opportunity/>
3. Anandan, R., Sipahimalani, R., Saini, S., Aryasomayajula, S., Smittinet, W. (2018). e-Conomy SEA 2018: Southeast Asia's internet economy hits an inflection point. Asia Pasific: Google-Temasek. Available at <https://www.thinkwithgoogle.com/intl/en-apac/future-of-marketing/digital-transformation/e-conomy-sea-2018-southeast-asias-internet-economy-hits-inflection-point/>
4. Rachman, Y. (6 July 2019). *Jalan terjal mengembangkan ekonomi digital*. AntaraNews. Available at <https://www.antaranews.com/berita/944414/jalan-terjal-mengembangkan-ekonomi-digital>
5. Peraturan Presiden Nomor: 74 Tahun 2017 Peta Jalan Sistem Perdagangan Nasional Berbasis Elektronik (Road Map E-commerce) Tahun 2017-2019. Lembaran Negara Republik Indonesia Tahun 2017 Nomor 176. Jakarta. Available at <https://peraturan.bpk.go.id/Home/Details/73107/perpres-no-74-tahun-2017>
6. U.S. Bureau of Economic Analysis. (Maret 2018). Digital Economy. Available at <https://www.bea.gov/data/special-topics/digital-economy>.
7. Mesenbourg, T. (2001). Measuring The Digital Economy. Available at <https://www.census.gov/content/dam/Census/library/working-papers/2001/econ/umdigital.pdf>
8. Laudon, K.,Traver, C. (2017). E-commerce 2016: Business, Technology, Society, 12th edition. Pearson. Available at <https://www.pearson.com/store/p/e-commerce-2016-business-technology-society/P100000775967>.
9. Wijaya, AP., Mariyah, S. (2019). Study of Consumer Price Index based on E-Commerce in Indonesia. Asia-Pasific Economic Statistics Week. UN ESCAP. Bangkok, Thailand. DOI: 10.13140/RG.2.2.15682.09924
10. Kestenbaum Richard. (26 April 2017). What Are Online Marketplaces and What Is Their Future? Forbes. Available at <https://www.forbes.com/sites/richardkestenbaum/2017/04/26/what-are-online-marketplaces-and-what-is-their-future/#563510603284>
11. Badan Pusat Statistik. (2019). Statistik E-Commerce 2019. Jakarta: BPS. Available at <https://www.bps.go.id/publication/2019/12/18/fd1e96b05342e479a83917c6/statistik-e-commerce-2019.html>
12. Mitchell, T. (2019). E-commerce and the Consumer Price Index: Measuring Inflation in a Digital Economy. Published by authority of the Minister responsible for Statistics Canada © Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2019. Available at <https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2019008-eng.htm>
13. Cavallo, A., Rigobon, R. (2016). The Billion Project: Using Online Prices for Measurement and Research. Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 151–178.
14. Badan Pusat Statistik. (2013). Buku Pedoman Survei Harga Konsumen. Jakarta: BPS.
15. Badan Pusat Statistik. (2020). Buku Pengolahan Indeks Harga Konsumen, Buku 5. Jakarta: BPS.
16. Patel, R., Patel, M. (2014). A Survey on Information Retrieval From Web Using Web Scraping Technique. IJIRT. Volume 1 Issue 6. ISSN : 2349-6002. International Journal of Innovative Research in Technology 1(6):95-100. Available at <http://ijirt.org/master/publishedpaper/IJIRT100512_PAPER.pdf>
17. International Monetary Fund. (2020). Consumer Price Index Manual Concepts and Methods. Available at <https://www.imf.org/~/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx>.
18. Badan Pusat Statistik. (2020). Diagram Timbang Indeks Harga Konsumen Hasil Survei Biaya Hidup 2018, Buku 2. Jakarta: BPS.Available at <https://www.bps.go.id/publication/2020/03/27/ae8a218c944f592cfbfc5937/diagram-timbang-indeks-harga-konsumen-hasil-sbh-2018--buku-2.html>
19. Pramana, S., Yuniarto, B., Kurniawan, R., Yordani, R., Lee, J., Amin, I., Satyaning P.P., N., Riyadi, Y. (2017). Big data for government policy: Potential implementations of bigdata for official statistics in Indonesia. Jakarta. International Workshop on Big Data and Information Security (IWBIS), pp. 17-21, doi: 10.1109/IWBIS.2017.8275097.
20. United Nations. (2009). Practical Guide to Producing Consumer Price Indices. Geneva: Switzerland. Available at <https://unece.org/fileadmin/DAM/stats/publications/Practical_Guide_to_Producing_CPI.pdf>
21. Tukey, J. (1977). Exploratory Data Analysis. Philippines: Addison-Wesley Company, Inc. Available at <http://www.ru.ac.bd/wp-content/uploads/sites/25/2019/03/102_05_01_Tukey-Exploratory-Data-Analysis-1977.pdf>
22. Salkind, N. J. (2010). Encyclopedia of research design (Vols. 1-0). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412961288

1. Muhammad Ghozy Al Haqqoni, BPS-Statistics Statistics Indonesia.

   ORCID: <https://orcid.org/0000-0001-9239-1544> [↑](#footnote-ref-1)
2. Corresponding Author: Setia Pramana, Politeknik Statistika STIS, Jakarta, Indonesia;

   E-mail: [setia.pramana@stis.ac.id](mailto:setia.pramana@stis.ac.id). ORCID: <https://orcid.org/0000-0002-8590-1451> [↑](#footnote-ref-2)