<table>
<thead>
<tr>
<th>Seminar Component</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name of author</strong></td>
</tr>
<tr>
<td>Alifa Putri Wijaya, S.Tr.Stat. ()</td>
</tr>
<tr>
<td>Siti Mariyah, S.S.T.,M.T()</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
</tr>
<tr>
<td>BPS – Statistics Indonesia- Katingan</td>
</tr>
<tr>
<td><strong>Contact address</strong></td>
</tr>
<tr>
<td>BPS – Statistics Indonesia- Katingan</td>
</tr>
<tr>
<td>Jl. A. Yani Kompleks Perkantoran Kereng Humbang Kasongan</td>
</tr>
<tr>
<td>Katingan, Kalimantan Tengah, 74411, Indonesia</td>
</tr>
<tr>
<td>Telp/Faks (0536) 4043603,</td>
</tr>
<tr>
<td>Mailbox : <a href="mailto:bps6209@bps.go.id">bps6209@bps.go.id</a></td>
</tr>
<tr>
<td><strong>Contact phone</strong></td>
</tr>
<tr>
<td>(+62) 812 2252 2319, (+62) 85 6814 3290</td>
</tr>
<tr>
<td><strong>Email</strong></td>
</tr>
<tr>
<td><a href="mailto:alifaputri.victory@gmail.com">alifaputri.victory@gmail.com</a> ; <a href="mailto:sitimariyah@stis.ac.id">sitimariyah@stis.ac.id</a></td>
</tr>
<tr>
<td><strong>Topic of seminar to which your abstract relates:</strong> Alternative data sources for production of official economic statistics and SDG indicators</td>
</tr>
<tr>
<td><strong>Title of paper:</strong> Study of Consumer Price Index based on E-Commerce in Indonesia</td>
</tr>
</tbody>
</table>
Abstract

The increasingly rapid development of e-commerce in Indonesia, this clearly has an influence on commodity demand and supply, which results in price changes. The Statistics of the Republic of Indonesia (BPS) captures information on changes in commodity prices through the Consumer Price Index (CPI). CPI is one of the important economic indicators that can provide information about the development of prices of goods/services paid by consumers. Changes in CPI describe the rate of increase or decrease in the level of goods and services, known as inflation. Inflation is one of the important indicators in measuring the poverty level of a country, which now the country is also trying to realize number 1 SDGs which are "No Poverty". But the CPI captured by BPS has not included e-commerce as a respondent. Therefore, this study aims to capture price changes in e-commerce to produce e-commerce-based CPI through the creation of commodity price web scraper and evaluate e-commerce CPI results with BPS CPI results in each commodity. There are 5 stages carried out by researchers to achieve these objectives, namely preparation of data extraction, making a web scraper and scheduling data extraction, cleaning and compiling data, processing results and analyzing results. The movement of average commodity prices in e-commerce is more volatile than the average movement of commodity prices in the city of Jakarta, whose respondents are traders in traditional markets, modern markets and shops / outlets. In addition, it can be generally seen that the average commodity prices in e-commerce are higher than the average commodity prices in Jakarta. The process of calculating the CPI used is what BPS uses with modifications. In its implementation there are various benefits such as price data that is more reliable, cost-effective and labor-intensive. In addition, there are also several obstacles to producing e-commerce-based CPI, such as changes in e-commerce structure and the absence of supporting information such as the quantity of goods sold and weighing. So that further research is needed to overcome these obstacles in order to improve the quality of CPI calculations then policies taken to reduce poverty levels are more appropriate.
I. Contents

I. Contents .................................................................................................................................................. 3
II. Introduction ............................................................................................................................................. 4
III. Methodology ......................................................................................................................................... 4
   A. E-Commerce Concept .......................................................................................................................... 5
   B. Consumer Price Index (CPI) Concept ................................................................................................. 5
   C. Limitations ................................................................................................................................. Error! Bookmark not defined.
   D. Research Steps ............................................................................................................................... 7
      1. Research Framework ..................................................................................................................... 7
      2. Data Collecting Method ............................................................................................................... 7
      3. Data Processing Method ............................................................................................................. 7
IV. Discussion ............................................................................................................................................ 8
   A. Analysis and Evaluation of Data Extraction Preparation Phase ...................................................... 14
   B. Analysis and Evaluation of Web Scraper Making and Data Extraction Scheduling ...................... 14
   C. Analysis and Evaluation of Cleaning and Compiling Data Phase .................................................. 18
   D. Analysis and Evaluation of Results Processing Phase ................................................................. 18
   E. Analysis and Evaluation of Average Prices of E-commerce Commodities with Average Commodity Prices in Jakarta City ..................................................................................... 18
   F. Analysis and Evaluation based on E-commerce-based CPI Results with BPS CPI .................. 20
V. Conclusions ........................................................................................................................................... 21
   A. Conclusion ....................................................................................................................................... 23
   B. Suggestion ...................................................................................................................................... 23
VI. References ............................................................................................................................................ 24
VII. Appendixes ......................................................................................................................................... 26
II. Introduction

The development of e-commerce has spread to various countries, including Indonesia. In Indonesia, e-commerce enthusiasts are increasing, both as sellers and as buyers. The Ministry of Communication and Information predicts that the volume of e-commerce businesses in Indonesia will reach US$ 130 billion in 2020 with a growth rate of around 40 percent per year. The increasing prediction of e-commerce business volume is also inseparable from the development of internet users in Indonesia. In 2014, the number of internet users in Indonesia was 88.1 million and this number increased quite rapidly in 2015 which reached 93.4 million people. From this statement, it is very possible that the emergence of e-commerce will affect the demand and supply of goods and services, which in turn affect market prices, especially at the consumer level. Thus, it can indirectly affect the Consumer Price Index (CPI) in Indonesia.

CPI is one of the important economic indicators that can provide information about change of goods/services prices paid by consumers. In order to get the changes, BPS collects commodities' prices manually through a direct field survey. Until now, price changes that have occurred in e-commerce have not been included in the CPI calculation. This can cause bias in the CPI figures produced by BPS RI. Meanwhile, the demand for BPS RI data which is more real-time and up to date is increasing.

The various e-commerce are developing in Indonesia as listed in the directory of members of the Indonesian E-Commerce Association (IDEA), for instance Hypermart, Klikmart, Gramedia, Bukupedia, Bhineka, Pusat Kosmetik, Sephora, Zalora, BerryBenka, Electronic City, Century Pharmacy, Babyzania, Stationary and Mothercare. E-commerce sells various goods / commodities with various brands and qualities. In order to capture information from such e-commerce, we built crawling and scraping programs. The programs visit e-commerce website and retrieve some information such as product names, product prices, products sold, etc. Then these information are analysed to identify the price changes occurred in a certain time period.

National Statistics Offices (NSOs) in several countries have begun conducting studies on the use of big data to retrieve information from e-commerce which can then be used to measure the CPI for instance New Zealand and Italy. Based on research entitled "Towards a Big Data CPI for New Zealand", New Zealand’s NSO conducted a study of the number of commodities that can be covered by e-commerce. The results of the study explain that there are several CPI commodities in New Zealand that cannot be covered by commodities in e-commerce. While research conducted in Italy with the title "Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP Compilation" explained that the use of web scrapping techniques can improve the quality of consumer price surveys in terms of time.

Therefore, to take price changes in e-commerce to produce e-commerce-based CPI, this study designed and implemented web scrapers that can retrieve commodity package prices used in compiling the CPI. In addition, this research was a preliminary study for BPS in producing CPI that utilizes e-commerce as a respondent, so this study also calculated CPI.
according to the calculation standards set by BPS and analysed CPI figures generated from e-commerce.

III. Methodology

A. E-Commerce Definition

According to Purbo (2001) e-commerce is a technology, application, and business process that connects companies, consumers, and communities through electronic transactions and trade in goods, services, and information electronically. According to the Indonesian E-Commerce Association (IDEA), E-commerce can be grouped as follows:

a. Classified Ads
   Is a classified advertisement site, where that site does not facilitate online transaction activities such as Kaskus and OLX.co.id.

b. Marketplace
   The business model where the website is concerned not only helps merchandise promotion, but also facilitates online money transactions for online traders, for instance Bukalapak, Elevania and Tokopedia.

c. Online Retail
   Online shop with its own website (domain) address where the seller has a stock of products / services and sells them online to buyers, for example Berrybenka, Bhinneka, Bukupedia, Electronic City and Zalora.

B. Concept of Consumer Price Index (CPI)

According to Buku Pedomen Survei Harga Konsumen 2013, CPI is one of the important economic indicators that can provide information on the development of prices of goods / services that are generally consumed by the public. The rate of change in the CPI, both experiencing inflation and deflation, is occurring, reflecting people’s purchasing power to meet their daily needs. The higher the inflation, the lower the value of money and the lower its purchasing power.

The CPI methodology generally carried out by BPS is as follows:

3.2.1 Consumer Price Data Collection

Sample design: the sample used in calculating CPI to choose cities, markets, respondents, commodities and quality, generally is done purposively with certain criteria "the most dominant". City Election: The selection of Consumer Price Survey (CPS) City was conducted in 82 cities (33 provincial capitals and 49 districts / cities). The selection of CPI cities is based on the relatively rapid level of development in the economy when compared to other cities.

Market Selection: Places to monitor consumer price data are traditional markets, supermarkets, and shops / outlets. Market criteria as a place to monitor consumer price data are relatively large and by the local community used as a comparison of prices, commodities, and the quality of other markets in the city concerned; located in the city area; various commodities can be found; many people shop there; long time shopping
crowds; continuation of Consumer Price (CP) data enumeration on the market must be guaranteed.

Selection of Respondents: CPS respondents are traders who sell goods / services for household needs in retail. The criteria for selecting traders as respondents are trading in harvesting, non-mobile places; various commodities traded; it is estimated that the continuity of CP enumeration is guaranteed; Traders are crowded with consumers; Easy to be interviewed, honest and friendly; and each commodity, 3-4 respondents were selected for data collection CP.

Selection of Commodity Types: BPS has commodity basket consisting of seven groups, 35 sub-groups, which have 875 types of commodities. The selection of commodities for each city is based on the results of the 2012 Life Cost Survey with the certain criteria. That criteria are the type of commodity has a percentage of consumption value to total household consumption ≥ 0.02 percent; the goods / services are consumed widely by the people of the city concerned; prices can be monitored continuously for a relatively long period of time. With these criteria, each CP city has a different number of selected commodities.

Selection of Quality or Brand: Each quality or brand commodity chosen should be a quality or brand that is much favoured by the people of the city concerned, quality/brand whose circulation is continuous and selected 3-4 qualities or brands for each commodity in cities in one province. Filling out the Questionnaire and Enumeration Time: There are ten types of questionnaires that are used with different objectives and frequency of enumeration.

3.2.2 Processing and Calculation of Consumer Price Statistics
In order to calculate the CPI, price data is needed from the results of enumeration in each city, on the market and respondents who have been determined. This price data then results in the city’s CPI and is subsequently combined into a combined CPI of 82 cities (National CPI). In general, the formula for calculating CPI by BPS uses the modified Laspeyres formula, as below:

\[
I_n = \frac{\sum_{i=1}^{k} P_{ni}P_{(n-1)i}Q_{oi}}{\sum_{i=1}^{k} P_{oi}Q_{oi}} \times 100
\]

where :
- \( I_n \) : nth period index
- \( P_{ni} \) : price of item type i, (n)th period
- \( P_{(n-1)i} \) : price of type i, (n-1)th period
- \( P_{(n-1)i}Q_{oi} \) : value of consumption of type of item, (n-1)th period
- \( P_{oi}Q_{oi} \) : value of consumption of type i on base year
- \( k \) : number of types of commodity package items

The processing stages as follows: calculating the Average Quality Price or Brand of Goods → calculating Relative Prices (RP) → calculating consumption value → calculating Index.
C. Research Steps

3.3.1 Research Framework

The research framework of this study started from the background of changes in transactions to online transactions through e-commerce. The transactions not only occur offline through traditional markets, modern markets, and stores but also occur online through e-commerce. Based on the interview conducted by BPS, it is known that currently the untapped e-commerce is the source of data used for calculating CPI. The change in CPI that occurs in a commodity has an influence on the achievement of the SDGs point 1. Therefore, study of CPI based on E-commerce was conducted. The methodology which used in this study and the full research framework, can be seen in Figure 2.

3.3.2 Data Collecting Method

There are several ways to collect data for this study, as follows:

1. Interview
   Interviews were conducted with staff of the Sub-Directorate of Consumer Price Statistics. Information about the flow of the CPI calculation process currently carried
out by BPS and information on the development of e-commerce based CPI being carried out by BPS were obtained from the interview.

2. Literature Study
The literature study was carried out by examining previous studies related to this study. The related research studied in this study pays attention to the use of theory, methodology and the results of the research.

3. Documents Review
The documents used were secondary data obtained from BPS. There are 4 documents used in this study:

- **CPI Basket Commodity**: This document contains the item code, group name, sub group name and also the type of commodity used in calculating CPI in 82 cities selected for the CPI calculation sample. This document is used to categorize products obtained from e-commerce.
- **Each Commodity Value for April 2018**: This consumption value was used to overcome the limitations in calculating CPI due to the unavailability of information regarding the quantity of goods sold.
- **Jakarta Price, Quality / Brand and Type of Commodity Data Jakarta April 2018 – June 2018**: This data was used as a comparison of the average price of each commodity, each month obtained from e-commerce.
- **National CPI for each commodity in May 2018 and June 2018**: This data is used as a comparison of CPI based on e-commerce.
- **2013 SHK Handbook (Buku Pedoman SHK 2013)**: This document was used to study the CPI calculation conducted by BPS.

4. Web scraping
The use of web scraping programs is to get secondary data in form of product information obtained from selected e-commerce. Product information in selected e-commerce were taken using the web scraper every week.

3.3.3 Data Processing Method

1. Preparation of Data Extraction
There are 3 stages, consist of:

   a. Determine the e-commerce site to be researched

      The selection of e-commerce used in this study applied a purposive sampling method with consideration of the selected e-commerce type online retail, the coverage of commodities obtained from e-commerce with the coverage of commodities in BPS, each type of commodity represented by 1 to 3 e-commerce. In addition, the selection of e-commerce was also based on the results of the 2016 E-commerce Award conducted by the Indonesian E-commerce Association (IDEA). The award assessment criteria are trust, convenience and e-commerce which are most often used in 8 cities (Greater Jakarta, Bandung, Semarang, Surabaya, Medan, Makassar, Denpasar and Palembang). So, with this method, the researchers obtained 14 e-commerce namely Hypermart, Klikmart, Gramedia, Bukupedia, Bhineka, Pusat Kosmetik, Sephora, Zalora, BerryBenka, Electronic City, Century Pharmacy, Babyzania, Stationary and Mothercare.

      In addition, the election of 14 e-commerce is based on how much coverage e-commerce can include BPS commodity basket. This study did not cover services...
commodity for several reasons. First, each service commodity has a different processing process, this can be seen in the 2013 consumer price survey guidebook. Second, according to the Global Online Consumer Report (2017), indicating that services commodity have not been maximally used in online transactions. This is indicated by goods commodity such as books, electronic goods and clothing (clothing) are items that are often purchased by online customers.

b. Record the type of item and the URL address of each e-commerce that will be scrape.

Every chosen e-commerce, has the type of item that it sells which contains various brands or qualities. For examples is for Electronic City (e-commerce) that sells various types of goods from refrigerators to lights. As can be seen in figure 2.

![Figure 2](image.png)

Figure 2. View of Website Category in Electronic City

From figure 2, when the air conditioner is selected, it will appear various qualities or brands sold as seen in figure 3.
From the information contained in Figure 4, researchers recorded the link / URL contained in e-commerce, then categorized the link based on the type of commodity contained in BPS commodity basket. Producing e-commerce commodity master files in the form of types of goods (group code, subgroup code, item type code), type of questionnaire (type of questionnaire, name of questionnaire, questionnaire description), type of e-commerce (e-commerce code, name e-commerce), and URL / Link. The master file is stored as excel csv. The e-commerce commodity master file in the next stage becomes an input file for the web scrapper.

c. Study the HTML structure of each e-commerce, learn the attributes or information can be retrieved.
As seen in figure 5, we can find out product information, product ID, old price (price before discount), new price (price after discount), and amount of discount. In order to find out what information is stored in HTML tags with what HTML structure, then do web page inspect, as in figure 6.

![Figure 6. HTML structure of Electronic City Website](image)

Figure 6 tells learn the HTML structure from every e-commerce. This stage was used as a reference in making web scraper and merging all attributes for preparation processing.

2. Development of Web Scraper and Scheduling of Data Extraction

![Figure 7. Flow of Web Scraper Work](image)

The web scraper formed in this study was designed to do crawling and scraping on e-commerce websites. Web scraper used Python 3 language with IDE Jupyter.
Notebook. The input needed to run the web scraper is an excel csv file that only contains an URL for each e-commerce. Once the web scraper is run, an excel output file is formed which contains data from attributes that can be taken on each e-commerce. Illustration as in figure 6. This activity is conducted once a week for each e-commerce. The one-week range used was from Monday - Sunday from April 2018 to June 2018. Data collection of this price is also what is done by BPS. Based on interviews conducted by BPS, it was explained that the reasons for taking time on certain commodities and monthly on other commodities were seen from how elastic a change in commodity prices and also because of limited labour and costs in data collection. Therefore, this activity took the frequency of the week because it did not require a lot of energy and also adjusts the ability of the web scraper.

3. Cleaning and Data Compiling
Data derived from the web scraper output could not be directly used for processing results. There were several e-commerce outputs from the web scraper whose attribute types were not ready to be processed. Consequently, it enforced cleaning data. Cleaning data was done in Microsoft Excel with the formulas and functions available in the software to data on each e-commerce can be ready to be processed. This cleaning process was focused on attributes such as product name, old price, new price, discount products, and e-commerce's URL.

The cleaning data process included:

a. Fixing stuffing
The contents of an e-commerce attribute were filled in according to the concepts and definitions. For example, when old price should be filled with the price before a discount, but the column also contains the new price which is the price after the discount, then it was necessary to do separation. With the purpose of the contents for the old price attribute must be filled by the price before the discount.

b. Fixing data types
The old price, new price, and product discount attributes must be numeric, while others have to be non-numeric. If the contents are not in accordance with these rules, fill out can be corrected.

c. Deleting out of stock data items.

![Diagram](image)

Figure 8. The document process is compiled
There were several e-commerce that provide information on product availability. If there is a product that is not available then the information from the product can be removed from the database. Furthermore, the excel files ready to be processed
are merged into one e-commerce data master file with illustrations like Figure 7. The file contains 15 attributes of all items sold and available on 14 selected e-commerce items and types of items based on BPS Commodity Basket from April 2018 to June 2018.

4. Result Processing
In general, the method used to process data from file output applied the method carried out by BPS in calculating CPI with several changes in assumptions. The processing method consists of several steps:

i. Sorting based on the type of item (commodity) and the name of e-commerce.

ii. Calculating the average type of goods in each week with the following formula:

\[
\bar{P}_{mi} = \frac{\sum_{j=1}^{J} P_{mi}^{j}}{J}
\]

where:
- \( \bar{P}_{mi} \) = Average price of week-m, commodity i
- \( P_{mi}^{j} \) = Week price, commodity i, brand j
- \( J \) = Number of Brands

iii. Calculating the average type of goods in each month with the following formula:

\[
\bar{P}_{ni} = \frac{\sum_{m=1}^{M} \bar{P}_{mi}}{M}
\]

where:
- \( \bar{P}_{ni} \) = Average price of period (n), commodity i
- \( M \) = Number of Weeks

iv. Calculating the relative price (RH) of commodities with the following formula:

\[
RH_{ni} = \frac{\bar{P}_{ni}}{\bar{P}_{(n-1)i}} \times 100
\]

where:
- \( RH_{ni} \) = Relative price of period (n), commodity i
- \( \bar{P}_{ni} \) = Average price of period (n), commodity i
- \( \bar{P}_{(n-1)i} \) = Average price of period (n-1), commodity i

v. Calculating the Consumption Value (NK) of the type of goods with the following formula:

\[
NK_{ni} = RH_{ni} \times NK_{(n-1)i} \times 100
\]

where:
- \( NK_{ni} \) = consumption value of period (n), commodity i
- \( RH_{ni} \) = relative price of period (n), commodity i
- \( NK_{(n-1)i} \) = consumption value of period (n), commodity i

Consumption value by April 2018 used the Consumption Value on BPS approach, because there was no availability of quantity information on commodities in e-commerce.

vi. Calculating the Consumer Price Index of the type of goods with the following formula:
where:

\[ N_{K_{ni}} = \text{consumption value of period (n), commodity i} \]

\[ N_{K_{0i}} = \text{basic consumption value of period (n), commodity i} \]

5. Result Analysis

Results processing yielded new data regarding the average price of each commodity each month (April, May, June) and e-commerce-based CPI for each commodity each month (May, June). The analysis used in this study was descriptive analysis. Descriptive analysis was used to analyse the advantages and challenges that occur in the preparation of e-commerce-based CPI. In addition, descriptive analysis with graphs was used to evaluate the average price of each commodity based on e-commerce with the average price of each commodity in Jakarta and also to evaluate the results of e-commerce-based CPI with CPI obtained by BPS for each commodity in the period May 2018 - June 2018.

IV. Discussion

A. Analysis and Evaluation of Data Extraction Preparation Phase

According to BPS commodity basket, this study succeeded in covering 6 groups, 23 sub-groups and 177 types of e-commerce commodities from 7 groups of 35 sub-groups and 875 types of commodities.
Figure 9 tells number of commodities each type which are covered by e-commerce and consumer price survey. The types of commodities covered by e-commerce are few enough. This is due to several reasons, namely as follows:

**Food and Cuisine**: Many kinds of dishes or foods are not yet possible to be sold and purchased online.

**Housing, Water, Electricity, Gas & Fuel**: there are still very few who sell and buy building materials online, and in this group there are many services such as rental houses, electricity rates, service items and wages for servants.

**Clothing**: actually this product is sold quite a lot in e-commerce, but the categorization of e-commerce is difficult enough to match the categories from BPS such as commodity basket shirts with long cotton shirts, short cotton shirts, batik shirts and others, while in e-commerce it only categorizes it as a shirt.

**Health**: there are a number of services not covered in e-commerce such as health services and physical care services

**Education, Recreation & Sports**: there are several commodity services such as course fees, equipment rental fees. In addition, there are still a few items that are still very little sold online such as stationery and magazines.

**Transport, Communication & Financial Services**: there is no commodity covered, because the commodity mostly contains service commodities and for commodities such as transportation equipment and spare parts not yet sold online.

The number of e-commerce commodities covered is still quite low compared to the number of commodities available in BPS Commodity Basket. This is because commodity e-commerce in this study does not involve commodity services

Table 1. List of Number of Commodities Taken in Every E-Commerce

<table>
<thead>
<tr>
<th>E-Commerce Name</th>
<th>Number of Commodity</th>
<th>E-Commerce Name</th>
<th>Number of Commodity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypermart</td>
<td>52 commodities</td>
<td>Babyzania</td>
<td>4 commodities</td>
</tr>
<tr>
<td>KlikMart</td>
<td>75 commodities</td>
<td>Apotek Century</td>
<td>10 commodities</td>
</tr>
<tr>
<td>Bhinneka</td>
<td>40 commodities</td>
<td>Pusat Kosmetik</td>
<td>5 commodities</td>
</tr>
<tr>
<td>Elektronik City</td>
<td>17 commodities</td>
<td>Sephora</td>
<td>2 commodities</td>
</tr>
<tr>
<td>Zalora</td>
<td>36 commodities</td>
<td>Stationary</td>
<td>6 commodities</td>
</tr>
<tr>
<td>BerryBenka</td>
<td>25 commodities</td>
<td>Gramedia</td>
<td>3 commodities</td>
</tr>
<tr>
<td>Mothercare</td>
<td>2 commodities</td>
<td>Bukupedia</td>
<td>3 commodities</td>
</tr>
</tbody>
</table>

This amount can still increase if further research will be carried out. Each type of item in this study comes from one to three e-commerce. Meanwhile, some of the conveniences that can be gained from implementing this stage when compared to the process of collecting price data
from BPS are: 1) cost saving because does not require the cost of procuring a questionnaire paper and does not need transportation costs to visit the respondent; 2) time saving: for the initial stages of doing this, it takes quite a lot of time. In this study, it took two months to gather this information. However, after this database is formed, it is enough to reduce the time to visit the respondent and only update the link / URL used.

Some obstacles and things that need to be considered at this stage are: 1) accuracy in categorizing a type of goods from e-commerce with BPS commodity basket. This research was done manually and did not use any system. Each link was entered one by one into the table then categorized the type of goods according to BPS commodity basket; 2) Links / URLs can change at any time: the process of checking and updating needs to be done before the link / URL is used as input on the web scraper.

B. Analysis and Evaluation of Web Scraper Making and Data Extraction Scheduling

The web scraper is created using python 3.6. This Web scraper script is run using the Jupyter Notebook IDE. This web scraper functions to retrieve all information / attributes contained in e-commerce and enter the data into database (excel).

The process that occurs in a web scraper as follows:

1. The web scraper opens an excel file that contains a list of links for each e-commerce
2. The web scraper opens the link and takes product attributes / information on the page
3. The web scraper enters that information in the data frame and converts it into an excel file for processing.

The main package used in the web scraper is Beautiful Soup and Selenium (only used on the Gramedia, Zalora and Sephora websites). The main difference between the Selenium package and the Beautiful Soup package is when Selenium is run, it will open a web browser (Mozilla Firefox). The used web scrapers were not wrapped on one application system because the HTML structure of e-commerce websites quite frequently changed. The web scrapers required a stable signal. There are 14 web-scraper for 14 selected e-commerce, which run every week. Weekly periods are selected for the following reasons: first, time used to run a web scraper was about 30 minutes to 3 hours. Due to this, a weekly period is chosen. Second, the need for a stable and fast signal that is difficult to fulfil.

The challenges and constraints obtained at the implementation of this stage were:

1. The structure of e-commerce websites often changed due to improvements to the appearance of the website or for the security of the website. As a result, the required attributes cannot be captured. Thus, resulting in the existence of several series of missing data not caught.
2. In addition to changing website structure, commodity links that had been stored in the master input file could no longer be up-to-date. So, it is necessary to update the master input file, if there was a change.

3. The availability of a stable internet signal to run the web scraper. Stable signal here also does not mean fast, too fast signal also made this web scraper stop working because even though it is fast in capturing information, this made the server administrator of e-commerce suspect that it was a crime. The result of an unstable signal was that we have to run the web scraper again.

4. Selenium package use has some problems, namely when the web scraper was run, the web scraper actually opened the link in the browser. Once the number of products taken was quite large, it affected the number of pages. If too many browsers were open on the computer then the computer became unresponsive and had the chance to experience a crash. The solution to the problem was to close the browser whose information has finished.

5. The use of regular expressions (regex), especially for cleaning data on search engines can slow the performance of the web scraper. The time needed to run the web scraper is longer and potentially the web scraper stops working, if the web scraper is made to produce data that is really clean. So, for some e-commerce sites, there is cleaning outside the web scraper.

The benefits obtained from the implementation of this stage when compared to the stages of data collection are as follows:

1. There is no non-response error

Compared to data collection from BPS, which need to visits respondents, it still has a very big opportunity for non-response with various reasons such as closed shops or because they really don’t want to be enumerated.

2. All information sold can be taken

Consumer price survey collect three-four qualities or brands of the type of goods that are most often sold. While with a web scraper, all the quality / brands of each type of item can be informed, so that with a lot of information it is expected that it can produce a more accurate CPI value.

3. Price data is more reliable

There is no intervention from the respondent when the enumeration was conducted. When BPS collects price data in the field, it is common for respondents to decrease or increase prices due to certain reasons that cause price data to decrease in credibility. In addition, not only the respondents made price changes, but they also happened to partners. Because of the high workload, partners can play price games. So that the price data obtained by credibility is reduced.

4. Energy Saving
In collecting BPS data, there are partners who are used to collect price data, so that a web scraper can reduce partner energy resulting in a reduction in partner accommodation costs.

C. Analysis and Evaluation of Cleaning and Compiling Data Phase

This stage was the longest stage of all stages that had been carried out. This stage produced data that was ready to be processed and calculated. Attributes that need to be considered in the cleaning and compiling process were ecommerce link, product name, old price, new price, product discount, update date, and availability. Cleaning was a process that aimed on data to be processed. While compiling is the process of combining all data obtained from e-commerce.

The obstacles and challenges that occur at this stage are as follows:

1. This stage requires precision and expertise to carry out the cleaning process because this process is still done manually using the excel formula.

2. When the website structure changes, it is possible that the contents taken can change, so that the cleaning process will also change

3. In calculating CPI, the same quality or brand from various selected respondents was calculated first. However, in this study the quality or similar brands from several e-commerce cannot be calculated.

D. Analysis and Evaluation of Results Processing Phase

As stated in the methodology, the calculation of average prices and CPI used the current method with some modifications. Calculation differences are in table 3.

Table 3. Differences in Data Processing between current method and this study

<table>
<thead>
<tr>
<th>No</th>
<th>Difference Location</th>
<th>Current method</th>
<th>E-Commerce</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Quality</td>
<td>The same quality is calculated as the average price first. After that, calculate the average commodity price</td>
<td>Cannot distinguish the same quality but it is located in a different e-commerce. Therefore, the average commodity price is calculated by calculating the average price of all qualities with similar types of commodities across e-commerce that sell these types of goods</td>
</tr>
<tr>
<td>2</td>
<td>Quality weighing</td>
<td>In conducting quality selection, BPS selects 3-4 quality of items that are often sold, and is given a</td>
<td>Because there is no information on goods and what quality is</td>
</tr>
</tbody>
</table>
weighting based on the results of 2012 SBH often sold, the weighter is omitted in this study.

<table>
<thead>
<tr>
<th>3</th>
<th>Consumption Value (NK)</th>
<th>The basic consumption values used to calculate the CPI using 2012 NK</th>
<th>To calculate the NK value for the current month, the BPS 201 NK value approach was carried out, and also the 2012 BPS Basic NK.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Weighing the city</td>
<td>The national CPI produced by BPS is a combined CPI of 82 cities. So that requires weighing the city</td>
<td>E-commerce cannot be found explicitly from where e-commerce originates. So that in the calculation of e-commerce-based National CPI, it does not use city weighter.</td>
</tr>
</tbody>
</table>

This study also showed that to calculate e-commerce-based CPI using the web scraping technique also requires additional information about the quantity of goods sold and also weighing. So that in further research, improvements can be made from the researcher. In addition, from the processing results there was a reduction in commodities from 177 commodities to 167 commodities, this was due to the influence of the cleaning process and changes in the structure of e-commerce. Details of the 167 commodities covered can be seen as follows.

![Changes in Initial Commodity Coverage](image)

**Figure 10. Changes in Initial Commodity Coverage**

From this processing phase, two dataset were obtained, namely the average e-commerce commodity prices April 2018 - June 2018 and e-commerce-based CPI May 2018 - June 2018.
E. Analysis and Evaluation of Average Prices of E-commerce Commodities with Average Commodity Prices in Jakarta City

The selection of Jakarta as a comparison of the average price generated by e-commerce is because Jakarta is the capital city of Indonesia which is the benchmark for economic growth in Indonesia. In addition, the City of Jakarta has a good development of information and technology and consumption patterns that are easily developed according to the growing trend. There are 396 commodities obtained from BPS related to commodity prices from April 2018 - June 2018. While there is an average price of 167 commodities collected from e-commerce. However, after further identification, there were only 122 types of similar commodities between Jakarta and e-commerce commodities. The coverage of these commodities can be seen in figure 10.

![Commodity Coverage for Average Price Analysis](image)

Figure 10. Commodity Coverage for Average Price Analysis

The results of evaluating the entire commodity are actually very interesting when discussing 122 existing commodities. However, in general it can be concluded that for the average movement of commodity prices in e-commerce more volatile than the average movement of commodity prices in the city of Jakarta, the respondents are traders in traditional markets, modern markets and shops / outlets. This is because the average price in the city of Jakarta has comparable quality in each of its respondents even though the quality that is successfully recorded is slightly less than in e-commerce. While in e-commerce even though the quality can be recorded more, but it does not have a quality comparison. Actually, this can be done by choosing similar qualities that are on some e-commerce, but this takes a long time.

In addition, it can be seen that most of the average commodity prices in e-commerce are higher than the average commodity prices in Jakarta. Some commodities that show these
phenomena are Men's Sports Shoes, Men's Sports Clothing, Balls, Tablet Computers, Flash Disks, Children's Bicycles, Cameras, HANDY-CAM, and others (can be seen in Appendix 3). This is also supported by research entitled Segmenting Online Shoppers: A Study of Consumers' Web Portal Selection Motivations For E-Shopping which proves that online market segmentation is enjoyed by middle and upper income consumers. So, for this reason, the average commodity prices in e-commerce are higher than in Jakarta. In addition, the average commodity prices in e-commerce are higher than in Jakarta, which can also be caused by very little interest in e-commerce for certain commodities. We can see this from clothing and beauty products whose average price in e-commerce is lower than the price in Jakarta. Like a phenomenon that can be seen on the internet, e-commerce that sells clothes is not only one or two but many, not only well-known brand companies that sell, but anyone can sell it. Competition is quite tight in e-commerce, making the price of these commodities lower. The results of the comparison of average prices can be seen in Appendix.

F. Analysis and Evaluation based on E-commerce-based CPI Results with BPS CPI

The unavailability of supporting data from e-commerce such as the quantity of commodities sold, weighing quality and weighing cities are the main limitations in calculating e-commerce based CPI. However, after going through a number of accounting processes and approaches, this research produced e-commerce based CPI. Although the researchers realized that the results produced could still be improved with more efficient techniques and approaches.

Of the 167 commodities whose average price can be calculated, only 163 commodities can be calculated for the CPI. The missing commodities are Mie Telor, Milk for Babies, Granulated Sugar and Racket. This is because information on these products only exists in April, so CPI calculation cannot be carried out. From 163 commodities, this study succeeded in forming e-commerce based CPI in May 2018 and June 2018. From these results it can be grouped into 9 parts, namely:

1. E-commerce CPI charts and BPS CPI Graphs are both rising
   There are 54 commodities that experience this phenomenon.
2. E-commerce CPI charts and BPS CPI charts are both decreasing.
   There were 13 commodities that experienced this phenomenon.
3. E-commerce CPI charts and BPS CPI charts were not experiencing changes
   There are 2 commodities that experience this phenomenon, namely the picture book and calculator
4. E-commerce CPI charts rise and the BPS CPI Graph decreases
   There are 13 commodities that experience this phenomenon
5. E-commerce CPI charts are up and the BPS CPI Graph has not changed
   There are 11 commodities that experience this phenomenon
6. E-commerce CPI chart decreases and BPS CPI Graph rises
   There are 54 commodities experiencing this phenomenon
7. The E-commerce CPI graph is decreasing and the BPS CPI Graph has not changed
   There are 11 commodities that experience this phenomenon
8. E-commerce CPI graphs have not changed and the BPS CPI Chart is up
   There are 6 commodities that experience this phenomenon
9. The E-commerce CPI graph has not changed and the BPS CPI Chart is declining
   There are no commodities that experience this phenomenon

For the most part, it can be seen that the e-commerce CPI has increased and decreased more with the BPS CPI when it experienced an increase and decrease. This is because the average e-commerce price is quite fluctuating. This can be seen in the attachment. This fluctuating e-commerce-based CPI changes, convincing us all that it is important for us to capture information originating from e-commerce so that the CPI measurement becomes more precise. With the right CPI measurement method, the measurement of inflation is better. Thus, all policies taken for the first SDGs (No poverty) can be more appropriately taken because the method of measuring inflation is better by entering information from e-commerce as a data source.

V. Revision

A. Volatility

Volatility measures variability, or dispersion about a central tendency — it is simply a measure of the degree of price movement in a stock, futures contract or any other market. According to Jeffry Bloem, there are 3 steps to calculate volatility, as follows:

1. Step 1 - Calculate Price Changes
   There are two generally accepted ways to calculate price changes. The first is the arithmetic method and the second is the logarithmic method. But, in this paper, we used Arithmetic Method with formula as follows
   $$\text{Arithmetic Price Change} = \frac{P_2 - P_1}{P_1} = \left(\frac{P_2}{P_1}\right) - 1$$

2. Step 2 – Calculate Standard Deviation
   Here the standard deviation with a 3-month time horizon calculates the average monthly price volatility.

3. Step 3 – Annualize Price Volatility
   The last step is to annualize the unit specific price volatility calculation from the previous step. To do this the standard deviation is multiplied by the square root of the number of observations.

Higher volatility indicates that the value of the price can be spread out over a larger range of values which eventually means that the value of the price can potentially move in either direction significantly over a short period of time. On the other hand, lower volatility indicates that the value of the price would not fluctuate much and will continue to remain stable over the period of time. Volatility price based on e-commerce can be compared with volatility price based on BPS, it can be seen in Appendixes.
B. More Explanation

Based on Figure 2. Research Framework, it explains that price in e-commerce can be a new data source for raising accuracy of CPI in a country. It not means that the old data source can not be used to calculate the CPI. The collaboration of all data source can make calculating of CPI more precise and accurate, remembering that trading (buying and selling) via online now became a new trending lifestyle. But, the new collaboration data source on calculating CPI need more research to do.

VI. Conclusions

A. Conclusion

Based on the research that I have done, it can be concluded as follows:

1. This study used the technique of web scraping, which uses a web scraper to collect product information on various types of goods from e-commerce, where things that need to be considered are internet signal stability and changes in e-commerce website structure.

2. The movement of average commodity prices in e-commerce is more volatile than the average movement of commodity prices in the city of Jakarta, where the respondents are traders in traditional markets, modern markets and shops / outlets. In addition, in general it can be seen that the average commodity price is at e-commerce is higher than the average commodity price in Jakarta

3. This research indicates that the ecommerce-based CPI calculation process can be done using a calculation approach conducted by BPS, but the unavailability of supporting information such as the quantity of goods sold frequently and quality weighing is an obstacle in this study, which is overcome by approaching BPS data with the 2012 base year.

4. Every step taken to produce e-commerce-based CPI has the advantages and challenges available. The benefits obtained are in the form of cost-effective and energy saving. If the data retrieval system until data processing can be more automated, then this will save time. Whereas the challenge is changing the ecommerce structure that cannot be known with certainty, internet stability, and unavailability of supporting information (quantity and weighing)

5. This fluctuating e-commerce-based CPI changes, convincing us all that it is important for us to capture information originating from e-commerce so that the CPI measurement becomes more precise. With the right CPI measurement method, the measurement of inflation is better. Thus, all policies taken for the first SDGs (No poverty) can be more appropriately taken because the method of measuring inflation is better by entering information from e-commerce as a data source.
B. Suggestion

There are several suggestions given by researchers so that this research can develop and can be used properly, namely as follows:

1. This research produces several topics that need to be developed so that e-commerce-based CPI calculations can be applied. These topics are as follows:
   - Development of web scraper for application, making it easier for users to use a web scraper. Other than that, Development of a web scraper to make automatic scheduling, so as to reduce errors that occur. Web Scraper users are also advised to understand programming well, so that when there is a change in structure it can be fixed immediately.
   - Similar quality checks on several e-commerce sites. How the system detects and categorizes quality is a topic that needs to be studied in future studies. Actually, there is a solution to this, namely the use of machine learning, but what methods are right to do and how efficient the use of the solution is, it needs to be examined further
   - Automation of CPI processing systems. Until now, BPS in processing data to produce CPI still uses Microsoft Excel-macro, while this study also still uses Excel because of the complexity of the CPI processing system.
   - The CPI calculation is a combination of e-commerce, traditional markets, modern markets and stores.

2. If BPS wants to use e-commerce as a new data source, the first thing to do is to establish good cooperation with e-commerce actors / sellers, thus facilitating BPS when it needs information from e-commerce. Such as information about the strength of goods that are often sold and also to invite e-commerce players to standardize the rules regarding the naming of goods, so that uniform categorization can facilitate the processing

3. E-commerce players need to support BPS's efforts to produce figures or data which can later be useful for e-commerce itself and the development of the trade climate.

VII. References


Bloem, Jeffrey. (2015). Calculating Price Volatility. Graduate Research Assistant, Department of Agriculture, Food, and Resource Economics, Michigan State University. Email: bloemjef@msu.edu. This document was originally created for use of the Food Security Group at MSU.


Giannini, R., Polidoro, F., & Virgillito, A. Web Scraping for Collecting Price Data: Are We Doing It Right?. Istat – Istituto Nazionale Di Statistica, Italy.


VIII. Appendixes

The Result of point E and F in Discussion Section can be seen in https://drive.google.com/open?id=1sTxfTCNSoKxOuAbY7HdGYpTJr69k5vkj

The Result of point A in Revision Section can be seen in https://drive.google.com/drive/folders/1hfKckM2ZZEatFyyF71znJpSUq4Kygzp2?usp=sharing