Online Motor Vehicle Sales Data for Supporting Policy in Manufacturing Sector

1st Satria Bagus Panuntun
Directorate of Statistical Analysis and Development
BPS Statistics Indonesia
Jakarta, Indonesia
satria.bagus@bps.go.id

2nd Khairunnisah
Directorate of Statistical Analysis and Development
BPS Statistics Indonesia
Jakarta, Indonesia
nisadzkiya@bps.go.id

3rd Dewi Krismawati
Directorate of Statistical Analysis and Development
BPS Statistics Indonesia
Jakarta, Indonesia
dewikrisma@bps.go.id

4th Setia Pramana
Politeknik Statistika STIS
Jakarta, Indonesia
setia.pramana@stis.ac.id

Abstract—Data and information on the income of the large and medium trade and the manufacturing sector are essential for the government to make policies. Currently, official statistics containing this information are still carried out conventionally. There has to be a new data source that can be used as a faster and more granular alternative reference. The goal of this research is to investigate a novel technique to generate data on vehicle sales in Indonesia from big data that can support and provide an overview of the manufacturing sector in real-time and may be used as a comparative or complementary to official statistics. This research uses a web scraping method from one of the largest vehicle advertiser sites in Indonesia. Vehicle sales data is collected weekly for the four vehicle types advertised in Indonesia using the HTML structure of the site. The Python programming language's Scrapy module is implemented. Data collection is carried out every week, and 358,451 vehicle advertisements have been collected from January 2019 to June 2021. The findings suggest that vehicle sales data from big data can be used as a comparative or complementary to official statistics, as well as supporting data in the manufacturing sector. By using web scraping techniques, indicators that usually require more time and cost can be done in real-time at a lower budget. This new approach is expected to improve the quality of official statistics in the manufacturing sector.

Index Terms—vehicle, sales, big data, mobil123

I. INTRODUCTION

The COVID-19 pandemic that has hit the world has changed the business processes of the economy, especially trade [1]–[3]. Mobility and direct interaction between humans have resulted in the spread of the COVID-19 virus getting faster. That has an impact on reducing trading activities that require human interaction. However, since the announcement of the first COVID-19 case in Indonesia in March 2020, the number of motor vehicle offers on Indonesian online platforms have increased rapidly. This transitional process illustrates a change in the trade business process by optimizing trade digitization from direct transactions where sellers and buyers meet to dealings with digital platform intermediaries.

Digital platforms such as marketplaces where sellers and buyers can make transactions without meeting each other are increasing during the COVID-19 pandemic. Bank Indonesia noted that the number of e-commerce transactions in the last five years has increased. Meanwhile, data collection on digital transactions has not been thoroughly carried out by the statistical office and generally still relies on conventional data collection, which takes a long time and costs a lot of money.
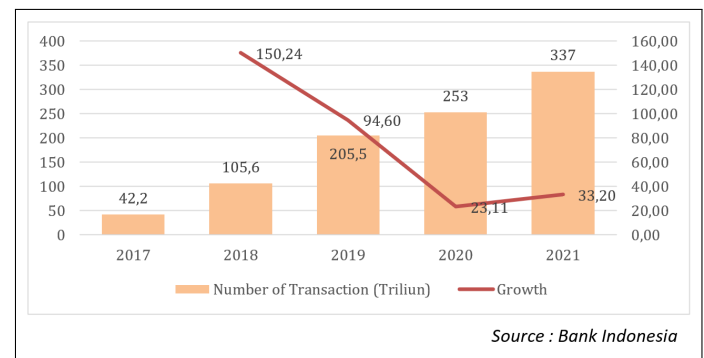

Fig. 1. Number and Growth Transaction Of Indonesia E-commerce.

Data on the digital marketplace platform can provide an overview of the demand and supply of goods or services. Those data include types of goods offered, prices offered, and the location of the supplier of goods. Besides that, the digital marketplace also provides reviews or recommendations systems from the buyer. The systems can build a trusted ecosystem to conduct online sales transactions than offline with more accessible agreements [4]. Some distributors only need to prepare motorcycles or cars already in demand by prospective buyers in online motor vehicle sales. That causes the distributor not to prepare the goods before a transaction is agreed upon. All those activities are recorded in digital applications and can be a new data resource for the official

statistic.

This digital transformation is increasing sales transactions without offline meetings in the pandemic era. For example, trading of luxury goods such as motorcycles and cars can continue without meeting offline by using online marketplace facilities or online platforms for selling cars and motorcycles. The shifting method gives companies advantages to knowing market conditions in real-time. Transaction sale agreements within the national scope occur not only in a few transactions but also in thousands of transactions on a daily, weekly, or even monthly basis. The number of online transactions will be overview by the number of cars or motorcycles offered online. Meanwhile, for official statistics, the data can be used to overview the income from the trade sector and the national manufacturing sector. Management of online marketplace data in real-time can describe the condition of the manufacturing sector with disruption conditions that occur in the market. It causes a decrease or increase in the demand and supply of goods and services. The increasing supply of motorcycles or cars on online platforms illustrates the manufacturing sector's performance, which is experiencing expansion. On the other hand, prices offered according to car and motorcycle brands describe the level of consumer preference in a certain period for cars and motorcycles offered in the market.

This study will collect vehicle sales data from online platforms by extracting vehicle advertising information from one of Indonesia's largest vehicle advertiser websites (Mobil123.com). Several studies on data collection using the web scraping method as a comparison or complementary data for official statistical data have been conducted [5]–[7]. In addition, UNESCAP also combines traditional and non-traditional methods (which is from big data) to obtain official statistics on consumer price indexes in several countries [8]. It is hoped that the collection of vehicle sales data from big data in this study can support and provide an overview of the manufacturing sector in real-time. The results can be used as comparison data or as a complement to official statistics. They can be an early warning to overcome the decline in performance in the manufacturing sector.

II. METHODOLOGY

A. Data Collection Method

In this research, data collection was carried out by collecting car and motorcycle (vehicle) sales data in Indonesia from the Mobil123 website. This site was chosen as a data source because it is one of Indonesia's most significant sources of information on car and motorcycle offers [9], [10]. The data collection process from the Mobil123 website can be separated into five parts: Preparation, Determining Web Scraping Technology, Vehicle Sales Data Collection, Weekly Scraping, Data Cleaning and Compilation (Figure 2).

1) Preparation: The first part in collecting data from the Mobil123 website is to determine a list of vehicle types and their URLs to be used as a base page for data collection. A list of vehicle types can be found on the homepage of the Mobil123 website. Figure 3 below is a display of the home
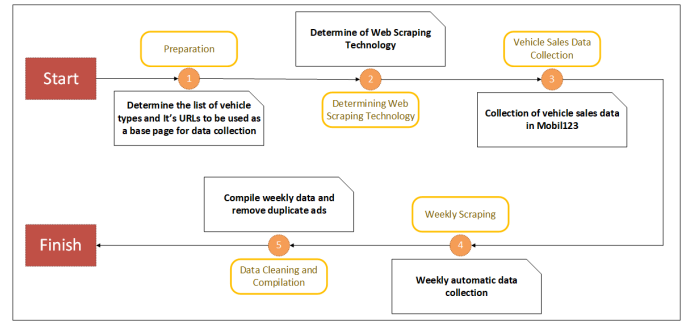

Fig. 2. Data Collection Workflow.

page which contains a list of the types of vehicles offered on the Mobil123 website. It can be seen that the types of vehicles are divided into four types, namely 'New car', 'Used Car', 'New Motorcycle', and 'Used Motorcycle'.
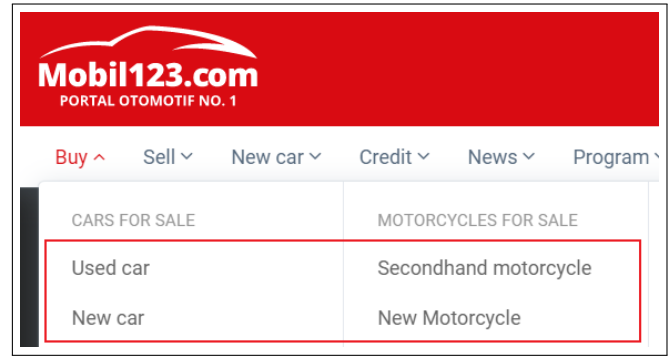

Fig. 3. List of Vehicle Types in Mobil123 Website.

Then the URL address will be determined from each vehicle type. The URL address structure of the vehicle type will be something like *https://www.mobil123.com/(vehicle type)-dijual/indonesia*. For example, the URL address for the New Car type page is *https://www.mobil123.com/ mobil-baru-dijual/indonesia*, as well as other types of vehicle. Next, all the URL addresses will be accessed one by one in the next step.

2) Determining Web Scraping Technology: After all URL addresses from each vehicle type in Mobil123 are determined, each URL address is accessed one by one. For example, Figure 4 below is the page displayed when a new car type URL is accessed. It can be seen in Figure 4 that there is a page that contains a list of new car ads in Indonesia. Then the URL addresses from the list of new car ads that appear on this page will be collected.

To collect the URL addresses from the list of new car ads, we need to access the HTML structure of this page. In this study, we use the Python Scrapy framework to complete the task. We use Scrapy because it can run asynchronously, which allows us to complete the vehicle sales data collection process faster than if we used Python's requests module [11]. Any information on a website page can be extracted using the Scrapy module by detecting the position of an element in the
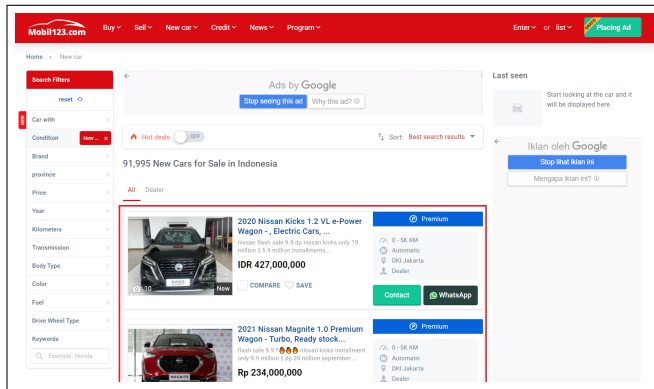
Fig. 4. List of Vehicle Ads in Mobil123 Website.

Fig. 5. An example of a page that contains detailed information regarding advertisements.

page's HTML structure [12]. For example the "2020 Nissan Kicks 1.2 VL e-Power Wagon" ad element as shown in Figure 4 is located at *h2 class = "listing__title"* in the HTML CSS structure of the page.

3) Vehicle Sales Data Collection: The focus of this study is to collect detailed information about each vehicle advertisement. To do this, we need to access each vehicle ad URL address that was collected in the previous step. For example, Figure 5 below is a page display that contains detailed information from the "2020 Nissan Kicks 1.2 VL e-Power Wagon" ad. On this page, there is some detailed information that can be extracted, including:

- Vehicle advertisement URL address
- Vehicle advertisement title
- Vehicle Price
- Advertisement date
- Seller type
- Seller location (country, province, and city)
- Vehicle type
- Vehicle brand
- Vehicle model
- Vehicle variant
- Vehicle year built
- Vehicle transmission
- Vehicle color

Later, all the information above from all vehicle ads will also be collected by marking the position of the HTML structure tag using the Scrapy framework in Python. For example, the position of the vehicle advertisement title is located at *h1 class = "listing__title"* tag in the HTML CSS structure of this page.

4) Weekly Scraping: This data collection is done once a week by collecting vehicle sales data from the previous week. This can be done because the Mobil123 site still stores their vehicle ads even though the advertisement has expired or sold. This weekly data collection is also carried out so that the workload of the scraper is not too heavy because it will only retrieve data for the past week.

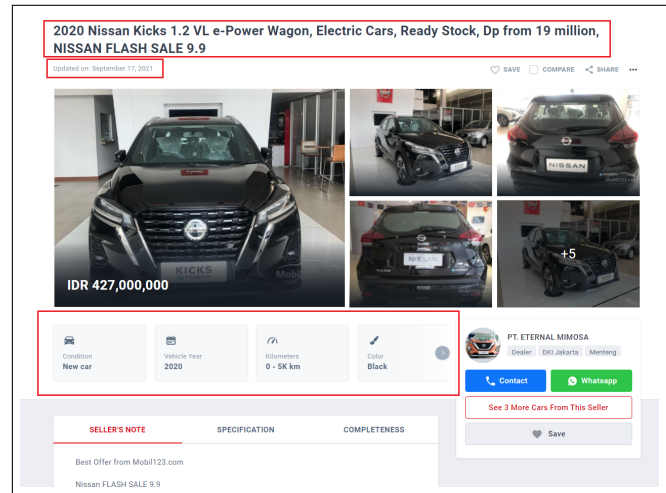*5) Data Cleaning and Compilation:* Before analyzing vehicle advertisement data collected as a result of the Mobil123 web scraping process, it must be cleansed. Cleaning includes the following steps:

- Changing the data type of the data field. Changing the data field is required so that we can properly examine the results.
- Removing duplication of data based on vehicle advertisement URL address

After the cleaning procedure is completed, the cleaned data from a certain day is saved in both the local database and the cloud. In a local database, data is stored in a MySQL formatted database, however in the cloud, it is stored in Big Query using SQL. We choose Big Query because it can manage data in real-time utilizing quick SQL-like queries.

B. Data Analysis Method

In this research, the analysis of vehicle sales data that has been collected is carried out in several steps. Data analysis begins with data visualization and descriptive data analysis.

1) Data Visualization: Data that has been collected will be visualized with the Microsoft PowerBI application to make data more representative and interactive. With this application, we can create many kinds of visualizations such as line charts, bar charts, pie charts, and even world map-shaped charts. For vehicle sales data, we visualize it into line charts and bar charts.

2) Data Descriptive Analysis: The tabulated and displayed data collection is then evaluated by analyzing its characteristics. These variables include, among other things, the number of vehicle commercials per month, the number of car advertisements per month per brand, and others.

III. RESULTS AND FINDINGS

A. Data Collection Results

The data collection method yielded a data set with 358,451 rows of Indonesian vehicle sales data. This data set is collected from all vehicle advertisements from January 2019 to June

2021. Table 1 below shows the amount of vehicle advertising data by vehicle type and year.

TABLE I
TOTAL DATA ON VEHICLE OFFERED FOR SALE IN INDONESIA
(2019 - 2021)

Type of Vehicle	Year		
	2019	2020	2021
New Car	51909	75107	26221
Used Car	29026	122172	50848
New Motorcycle	31	72	66
Used Motorcycle	534	2107	358

Based on table 1, more cars have been offered than motorcycles in the last three years on online platforms. Almost 99 percent of the vehicles advertised on digital platforms are cars. Meanwhile, at the beginning of the development online vehicle sales platform, 64 percent of the car advertisements offered were new cars.

Increasingly developing online motor vehicle sales platforms, used car advertisements are more advertised than new cars. It was recorded that in 2020, 62 percent of the advertising offered were used cars than new cars, and similarly, in 2021, more used cars were offered than new cars. The same thing is also seen in motorcycles or two-wheeled vehicles that advertise more used than new ones. For further discussion will relate to car supply data in describing the manufacturing sector in Indonesia.

B. Data Visualization and Analysis Results

Data visualization can be performed based on data on new and used car advertisements obtained from online sales platforms, as shown in Figure 6. Figure 6 illustrates the development number of new and used car advertisements by month and year advertised. In addition, you can see the monthly growth of the advertised car. Generally used cars are broadcast more than new cars. However, in March and April 2020, new cars were advertised as bigger than used cars.

The conditions in March and April 2020 were the initial conditions of the pandemic in Indonesia, and the number of new cars offered increased without being accompanied by purchases. This supply condition illustrates the declining performance of the manufacturing sector in Indonesia at the beginning of the pandemic. Meanwhile, during the pandemic, new car advertisements showed a recovery in early 2021 by increasing in June with government intervention in the policy of buying new cars.

Development of manufacturing sector apart from the number of car advertisements, we can also see the proportion of car brands offered. The highest number of car brand advertisements in a period illustrates the choice preferences that are in demand by the market at that time. Figure 7 shows that the car market in Indonesia is in great order for the purchase of cars for the Toyota and Honda brands. This illustrates that new competitors should survive in the Indonesian car market to have abilities such as Toyota or Honda brand cars. In addition,
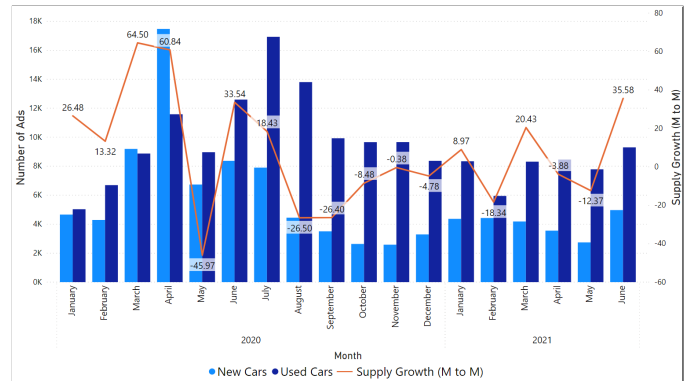

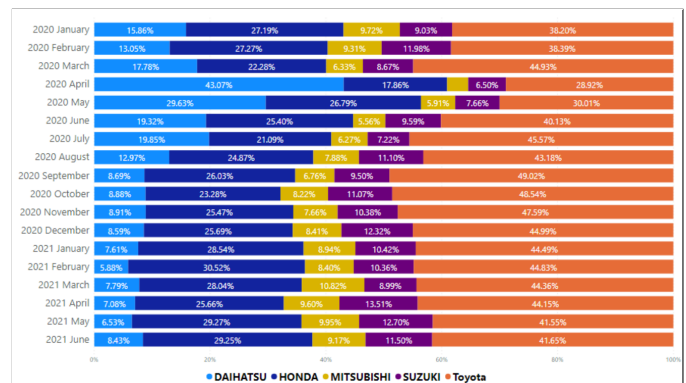Fig. 6. Number of Car Advertisements by Month.

Fig. 7. Number of Car Advertisements by Month and Car Brand

we can see price competition between car brands to survive in the Indonesian car market.
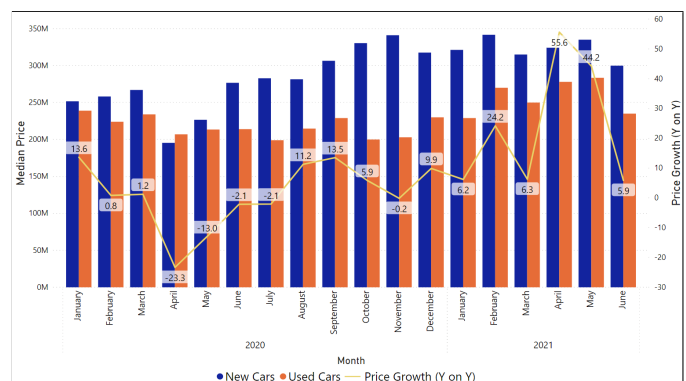

Fig. 8. Trend Median Price of Car advertisements by month.

Although the number of new cars offered is less than used cars, the median price of new cars advertised is still higher. However, at the start of the COVID-19 pandemic in April, the median price of new cars offered was lower than the price of used cars. That illustrates the slow-down market at the same time the median price of new cars is lower than used cars, with the number of new cars being offered more than used cars.

IV. Discussion

Based on the car sales data acquired from big data sources (web scraping from the Mobil123 website), these results can be stated to represent the condition of motor vehicle sales in Indonesia. Although the current results cannot be compared to official statistics, the existing data sets can be used as complement to official data and can be used as an illustration of the income of the large and medium trade sector and the manufacturing sector nationally. For example, from the results of data visualization and analysis, it can be seen that there was an increase in vehicle supply that was not accompanied by purchases in Indonesia in March and April 2020. This illustrates that there was a decline in the performance of the manufacturing sector in Indonesia at the beginning of the pandemic. Data consumers can receive information about the number of vehicle sales in Indonesia more rapidly using the data that has been collected (almost real-time). So it can be said that it is feasible to use big data to compare or complement official statistics in the manufacturing sector. By using web scraping techniques, indicators that usually require more time and cost can be done in real-time at a lower budget. This new approach will increase the quality of official statistics in the manufacturing sector.
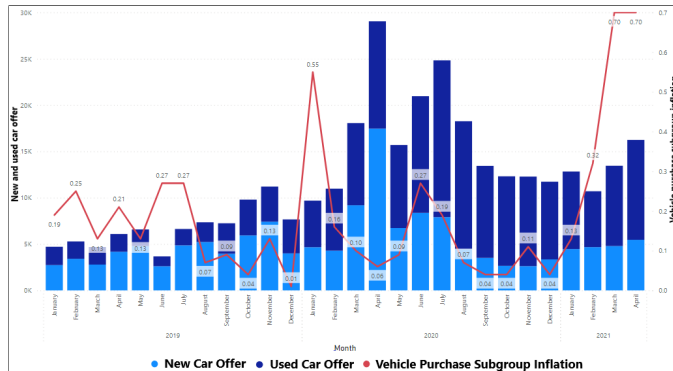

Fig. 9. Inflation of Motor Vehicle Purchase Sub-group and Number of Car Offers.

Figure 9 shows the data from the scrapping of car deals compared to the inflation data for the subgroup of vehicle purchases from official statistics, showing a relationship. During the pandemic in April 2020, the supply of cars increased from normal conditions in line with the decreasing rate of inflation for vehicle purchases. That illustrates declining of car trading, so policies are needed to overcome it.

Meanwhile, when compared to the Purchasing manager index, the same pattern occurred in April 2020 (Figure 10). The decreasing Purchasing Manager Index (PMI) in April 2020 reflects the manufacturing industry's declining optimism with the increasing level of car supply. This condition illustrates many vehicles produce without a purchaser. Meanwhile, when compared with the acquisition of PPnBM tax by month with the offer price of a car, it can be shown in Figure 11 when the offer price of a new car increases, it will reduce the purchase of new cars and also reduces the PPnBM tax revenue. This
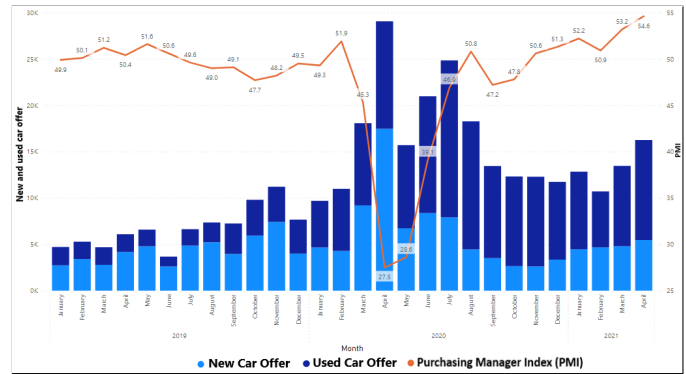

Fig. 10. Monthly PMI of manufacturing and Number supply of New and Used Car on Indonesian online platform.

condition can be an early warning system for policy in the manufacturing sector. In the future, this study can be based on developing a real-time dashboard for manufacturing policy.
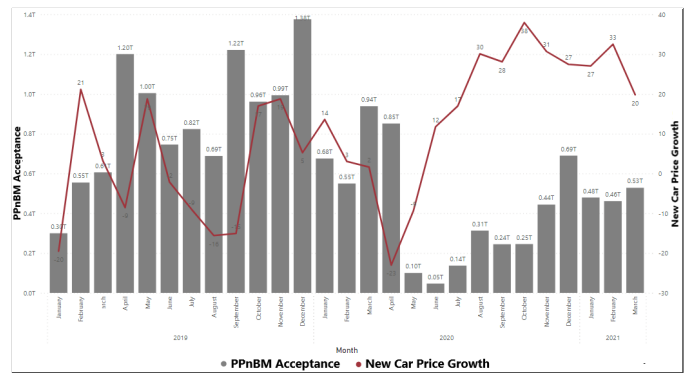

Fig. 11. Domestic PPnBM Monthly Revenue and Monthly Growth in New Car Prices.

V. Conclusion

The vehicle sales data set collected from big data using the web scraping method can be used to complement official data released by a statistics office. The kind of released data is the GDP manufacturing sector. In addition, this data set can also be accessed by users more quickly. This study indicates that there are 358,451 vehicles offered in Indonesia from January 2019 to June 2021 on the Mobil123 website, with cars as the most widely provided vehicle type. The results of this study also found a decline in the manufacturing sector's performance in Indonesia at the beginning of the pandemic in March and April 2020 was an increase in the motor vehicle offers that were not accompanied by purchases. These results can be used to make some policies to increase the demand for motor vehicles, like decreasing sales tax. Indonesia has decided to decrease the luxury goods tax for motor vehicles from March until December 2021 to increase demand for motor vehicles.

References

[1] P. E. Giordani and F. Rullani, "The Digital Revolution and Covid-19," SSRN Electronic Journal, 2020, doi: 10.2139/ssrn.3725851.

[2] R. Y. Kim, "The Impact of COVID-19 on Consumers: Preparing for Digital Sales," IEEE Engineering Management Review, vol. PP, p. 1, Apr. 2020, doi: 10.1109/EMR.2020.2990115.

[3] U. Bustaman, Khairunnisah, and W. A. Sofa, "Measuring the Signals of Indonesia's Economic Recovery," in Big Data Study: Signals for Indonesia's Recovery from the COVID-19 Pandemic, S. Pramana and D. N. Larasati, Eds. Jakarta: BPS Statistics Indonesia, 2021, pp. 43–62.

[4] S.-H. Liao, R. Widowati, and Y.-C. Hsieh, "Investigating online social media users' behaviors for social commerce recommendations," Technology in Society, vol. 66, Aug. 2021, doi: 10.1016/j.techsoc.2021.101655.

[5] S. Pramana et al., "Big data for government policy: Potential implementations of bigdata for official statistics in Indonesia," in 2017 International Workshop on Big Data and Information Security (IWBIS), 2017, pp. 17–21. doi: 10.1109/IWBIS.2017.8275097.

[6] Y. Adhinugroho et al., "Development of online travel Web scraping for tourism statistics in Indonesia," Information Research: an international electronic journal, vol. 25, no. 4, 2020, doi: 10.47989/irpaper885.

[7] S. B. Panuntun and S. Pramana, "Development of Automated Flight Data Collection System for Air Transportation Statistics," Journal of Physics: Conference Series, vol. 1863, no. 1, p. 12020, Mar. 2021, doi: 10.1088/1742-6596/1863/1/012020.

[8] UNESCAP E Statistics Division, "Incorporating Non-traditional Data Source into Official Statistics - The case of Consumer Prices Indexes: Lessons and experiences from Australia, Japan and New Zealand," 2020.

[9] Similarweb. Top Websites Ranking. Accessed on July 20, 2021, from https://www.similarweb.com/top-websites/indonesia/category/vehicles/vehicles/.

[10] Mobil123. About Us. Accessed on July 20, 2021, from https://www.mobil123.com/bantuan#tentang-mobil123.

[11] D. N. Myers and J. W. McGuffee, "Choosing Scrapy," 2015.

[12] Chaves A, "Scrapy Documentation Release 2.5.0 Scrapy developers," 2021. Accessed: Jul. 19, 2021. [Online]. Available: https://docs.scrapy.org/en/latest.