

# **HEDONIC PRICE INDEX FOR MOBILE PHONE BASED ON E-COMMERCE DATA**

**TRI LISTIANINGRUM**



**APPLIED STATISTICS  
GRADUATE SCHOOL  
IPB UNIVERSITY  
BOGOR  
2021**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



**@Hak cipta milik IPB University**

**IPB University**



**IPB University**  
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

## STATEMENT ABOUT THESIS, INFORMATION SOURCES, AND COPYRIGHT ASSIGNMENT

I hereby declare that the thesis entitled “Hedonic Price Index for Mobile Phone Based on E-Commerce Data” is my work with the direction of the supervising committee and has not been submitted in any form to any college. Resources originating or quoted from works published and unpublished from other writers have been mentioned in the text and listed in the Bibliography at the end of this thesis.

I hereby assign the copyright of my writing to the IPB University.

Bogor, January 2021

Tri Listianingrum  
NIM. G152184524

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## RINGKASAN

TRI LISTIANINGRUM. Indeks Harga Hedonis untuk Telepon Seluler Berdasarkan Data *E-Commerce*. Dibimbing oleh FARIT MOCHAMAD AFENDI dan BAGUS SARTONO.

*Big data* memberikan peluang untuk membangun kumpulan data yang sesuai dengan kebutuhan penelitian. Salah satu perhatian global mengenai pemanfaatan *big data* adalah dalam penggunaannya untuk mengukur indeks harga. Seiring dengan perkembangan situs *e-commerce* yang pesat, banyak informasi mengenai harga dan spesifikasi produk yang tersedia di internet. Informasi ini memungkinkan peneliti untuk menghitung indeks harga hedonis untuk produk yang mengalami perubahan kualitas yang cepat seperti telepon seluler.

Salah satu metode untuk mengumpulkan *big data* adalah dengan menerapkan pengumpulan data skala besar di web dengan proses *scraping* yang diautomatisasi. Pada penelitian ini dibuat *web scraper* dengan menggunakan bahasa *python* untuk mengekstrak informasi tentang produk, harga dan karakteristik dari berbagai *e-commerce* yang ditampilkan di website IPPrice.co.id. Ekstraksi data dilakukan setiap minggu dari Januari hingga Juni 2020 dan mengumpulkan lebih dari 34.000 amatan secara total. Sebuah database yang berisi spesifikasi lengkap dari 458 jenis telepon seluler dikembangkan berdasarkan informasi dari gsmarena.com untuk memvalidasi data hasil *scraping*.

Metode yang digunakan untuk menghitung indeks harga hedonis adalah *double imputation method*. Salah satu langkah dalam metode ini adalah membuat model regresi hedonis untuk dataset yang terdiri dari dua periode yang berurutan. Karena adanya pencilan dan nilai *leverage* yang tinggi, regresi kekar dengan estimasi MM (*Multi-Stage Method*) digunakan untuk mendapatkan estimasi parameter yang andal dan efisien. Tidak adanya informasi mengenai kuantitas produk sebagai dasar pembobotan membuat peneliti melakukan beberapa alternatif skenario pembobotan. Yang pertama adalah pembobotan berdasarkan tahun rilis telepon seluler, yang kedua adalah dengan menggunakan pangsa pasar sebagai *proxy* untuk pembobotan, dan yang ketiga adalah kombinasi keduanya.

Hasil penelitian menunjukkan bahwa pergerakan harga berdasarkan transaksi daring yang ditangkap dalam penelitian ini tidak mencerminkan transaksi luring yang tercermin dalam Indeks Harga Konsumen (IHK) yang dihasilkan oleh Badan Pusat Statistik (BPS). Indeks mingguan yang dihasilkan memiliki fluktuasi yang lebih besar daripada IHK. Pembobotan yang didasarkan pada tahun rilis dan pangsa pasar menghasilkan indeks yang paling stabil.

**Kata kunci:** Indeks Harga Daring, Regresi Hedonis, *Web Harvesting*

## SUMMARY

TRI LISTIANINGRUM. Hedonic Price Index for Mobile Phone Based on E-Commerce Data. Under supervision from FARIT MOCHAMAD AFENDI and BAGUS SARTONO.

Big data provides an opportunity to build datasets that fit specific requirements for research. One of the global concerns about big data is how to utilize it for measuring price index. Due to the proliferation of e-commerce sites, a vast amount of prices information followed by the product specifications are available on the internet. This information would allow researchers to calculate hedonic indices for the products undergo rapid quality change like a mobile phone.

A method to gather big data is by implementing a large-scale data collection on the web with an automated scraping process. In this study, a web scraper was built using python language to extract information about product, price and characteristics from various e-commerce displayed on the IPrice.co.id website. The data extraction was conducted weekly from January to June 2020 and gathered more than 34,000 records in total. A database containing comprehensive specifications of 458 types of mobile phones was developed based on information from gsmarena.com to validate the scraped data.

The method used for calculating the hedonic price index is the double imputation method. One of the steps in this method is to construct a hedonic regression model for datasets consists of two consecutive periods. Due to the presence of outliers and leverage points, robust regression with MM (Multi-Stage Method) estimation was used to get a reliable and efficient estimate of the parameters. Due to the lack of information on the product quantity for weighting, several alternative weighting scenarios were taken. The first is weighting based on the release year mobile phone, the second is by using market share as a proxy for weighting, and the third is the combination of both.

The result shows that price movement based on online transactions captured in this study do not reflects the one from offline transactions captured in the Consumer Price Index (CPI) produced by Badan Pusat Statistik (BPS). The weekly indices have a wider fluctuation than the CPI. Weighting based on both release year of mobile phone and market share produces the most stabilized index.

**Keywords:** Hedonic Regression, Online Price Index, Web Harvesting

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Copyright owned by IPB, 2021  
Copyright Act protected by Law

*Prohibited from quoting part or all of this paper without including or citing sources. Citations only for educational purposes, research, writing papers, preparing reports, writing criticism, or review a matter, and the citations do not harm the interest to IPB.*

*Prohibited from announcing and reproducing part or all of the papers in any form without permission from IPB.*

# **HEDONIC PRICE INDEX FOR MOBILE PHONE BASED ON E-COMMERCE DATA**

**TRI LISTIANINGRUM**

Thesis  
as one of the requirements to obtain  
Master of Science  
in  
Applied Statistics Program

**APPLIED STATISTICS  
GRADUATE SCHOOL  
IPB UNIVERSITY  
BOGOR  
2021**

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



*@Hak cipta milik IPB University*

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Examiner on Thesis Examination: Dr. Ir. Erfiani, M.Si.





Thesis Title : Hedonic Price Index for Mobile Phone Based on E-Commerce Data  
Name : Tri Listianingrum  
Student ID : G152184524

Approved by

Supervisor:

Dr. Farit Mochamad Afendi, S.Si., M.Si.

---

Co-Supervisor:

Dr. Bagus Sartono, S.Si., M.Si.

---

Acknowledged by

Head of Applied Statistics Program:

Dr. Kusman Sadik, S.Si., M.Si.

NIP 19690912 199702 1 001

---

Dean of Graduate School

Prof. Dr. Ir. Anas Miftah Fauzi, M.Eng.

NIP 19600419 198503 1 002

---

Examination Date:  
January 20<sup>th</sup> 2021

Graduation Date:  
January 29<sup>th</sup> 2021



## PREFACE

Alhamdulillahirabbil‘alamin, praise is mere to the Almighty Allah SWT for the gracious mercy and tremendous blessing that enabled me to finish this study in the middle of the Covid-19 pandemic. This study conducted from July 2019 to January 2021 is a thesis titled “Hedonic Price Index for Mobile Phone Based on E-Commerce Data”.

I would never have been accomplished this thesis without the support of many people. Therefore, I wish to show my appreciation to my research supervisors: Dr. Farit Mochamad Afendi, S.Si., M.Si. and Dr. Bagus Sartono, S.Si., M.Si. for their valuable assistance in developing this thesis. I want to thank my husband and son for their patience and understanding, in times when this woman is switching to college student mode, and for always blesses me with statistically significant joy in life. I wish to extend my special thanks to college mates from BPS STT 2018 for the companion and collaboration throughout the journey.

I hope that the result of this thesis could be an added value for science, especially in the development of price index methodology in Indonesia. Above all, hopefully, the completion of this study could make me a wiser and humbler person.

Bogor, January 2021

*Tri Listianingrum*

## TABLE OF CONTENTS

TABLE OF CONTENTS	vii
LIST OF TABLES	viii
LIST OF FIGURES	viii
LIST OF APPENDICES	ix
I INTRODUCTION	1
1.1 Background	1
1.2 Objectives	3
1.3 Benefit	3
1.4 Limitations	3
II METHODS	4
2.1 Research Framework	4
2.2 Data Scraping	5
2.3 Data Cleaning	8
2.4 Hedonic regression	11
2.5 Double imputation method	13
III RESULT AND DISCUSSION	16
3.1 Data Exploration	16
3.2 Hedonic Regression Analysis	19
3.3 Double Imputation Index	26
IV CONCLUSION AND RECOMMENDATION	33
BIBLIOGRAPHY	34
APPENDICES	37
BIOGRAPHY	45

Hak Cipta Dilindungi Undang-undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.  
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## LIST OF TABLES

1	Example of entries in the scraped data	8
2	Example of entries in the specification database	9
3	Example of entries in Product and Type attributes	9
4	The formation of the ID attribute	10
5	Example of entries in the final dataset	10
6	The estimates of hedonic model coefficients based on OLS method	22
7	The estimates of hedonic model coefficients based on OLS vs MM estimation methods	24
8	Percentages of matched and unmatched observations in the dataset from two consecutive periods	26
9	Weighting scenario 1	28
10	Weighting scenario 2	30

## LIST OF FIGURES

1	Research framework	4
2	Hedonic price index for CPI compilation	5
3	Data scraping process	5
4	Indonesia's market share for mobile phone from 1 <sup>st</sup> quarter of 2019 to 1 <sup>st</sup> quarter of 2020	7
5	Illustration of the samples from period t-1 and t	13
6	Flowchart of the double imputation method	15
7	Amounts of weekly scraped data	16
8	Number of observations per week in the final dataset	17
9	Boxplot of mobile phone price per week (thousand rupiah)	17
10	The series of means for price, internal storage, RAM and main camera from January to June 2020	18
11	The comparison between price and Ln price as response variable	19
12	The trend of battery capacity from 2011 to 2020 (mAh)	20
13	The normal qq plot of residuals and the influence plot of the OLS hedonic model	23
14	The comparison of hedonic price index based on OLS and MM estimation method	25
15	The comparison between the double imputation and matched model indices	27
16	Number of weekly observations by year release of the mobile phone	28
17	The comparison of hedonic price indices of the mobile phone between the use of weighting scenario 1 and without weighting in double imputation method	29
18	Number of weekly observations by brand	29
19	The comparison of hedonic price indices of the mobile phone between the use of weighting scenario 2 and without weighting in double imputation method	30

20	The comparison of hedonic price indices of the mobile phone between the use of all weighting scenario and without weighting in double imputation method	31
21	Monthly double imputation index and mobile phone price index from CPI with January 2020 as base period	32

## LIST OF APPENDICES

1	Structure of the scraped data	38
2	Structure of the specification database	39
3	Structure of the final dataset	40
4	Weekly index for the matched model, double imputation, double imputation with weighting scenario 1 (w1), double imputation with weighting scenario 2 (w2) and double imputation with weighting scenario 3 (w3) with week 1 2020 as base period	42
5	Test for equality of variance between unweighted double imputation index and double imputation index with weighting scenario 3	43
6	Monthly index for double imputation method with weighting scenario 3 and the index for mobile phone in the CPI with January 2020 as base period	44

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



*@Hak cipta milik IPB University*

**IPB University**



**IPB University**  
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.