

Scraping Indonesian E-commerce Data to Help Customer's Decision Making

Tugas Akhir

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai
Derajat Sarjana Komputer**



Khairuman Indrawan

210711096

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2025**

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

SCRAPING INDONESIAN E-COMMERCE DATA TO HELP CUSTOMER'S DECISION MAKING

yang disusun oleh

Khairuman Indrawan

210711096

dinyatakan telah memenuhi syarat pada tanggal 23 Juli 2025

Dosen Pembimbing 1	: Th. Adi Purnomo Sidhi, S.T., M.T.	Keterangan Telah Menyetujui
Tim Penguji		
Penguji 1	: Th. Adi Purnomo Sidhi, S.T., M.T.	Telah Menyetujui
Penguji 2	: Stephanie Pamela Adithama, S.T., M.T.	Telah Menyetujui
Penguji 3	: Martinus Maslim, S.T., M.T.	Telah Menyetujui

Yogyakarta, 23 Juli 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

Acknowledgment

I would like to take this opportunity to express my profound gratitude to all those who have supported and contributed to the successful completion of this thesis. This work would not have been possible without the guidance, encouragement, and assistance I received from many individuals throughout this academic journey.

First and foremost, I would like to extend my sincere thanks and deepest appreciation to my supervisor, Mr. Zhu Xiao Rui, for his exceptional mentorship, continuous guidance, and unwavering support throughout the duration of this project. His insightful suggestions, valuable feedback, and encouragement have played a crucial role in shaping the direction and quality of this work. His patience and dedication to excellence have been truly inspiring.

I am also deeply grateful to Mr. Li Qing and Mr. Wu Hai Yong for their guidance, encouragement, and constructive input. Their expertise and willingness to support me in overcoming challenges during the research and development process were of great value and are sincerely appreciated.

I would also like to acknowledge the invaluable support of my family and close friends. Their understanding, motivation, and constant encouragement provided me with the strength and confidence to persevere, especially during the more challenging phases of this work. I am especially thankful to my family for their unconditional love and belief in my abilities.

Finally, I would like to extend my appreciation to all those, named and unnamed, who have contributed in any way to the completion of this thesis. Your kindness, assistance, and encouragement have made a lasting impression, and I am truly grateful for your presence along this journey.

Abstract

The rapid expansion of e-commerce platforms in Indonesia has made it increasingly difficult for consumers to compare product information across multiple websites. This research introduces a web application designed to help users make more informed purchasing decisions by efficiently aggregating and analysing product data. The application uses APIs and web scraping techniques to collect relevant data from various Indonesia e-commerce websites, which is then exported into a DataFrame format for further processing. Machine learning algorithms are applied to this data to generate product recommendations. These recommendations are ranked based on criteria such as price and sales performance. By streamlining the comparison process and presenting a ranked list of top products, the system enhances the user's e-commerce experience and reduces decision fatigue.

Keywords: API, E-commerce, Machine Learning, Product Comparison, Recommendation System, Web Scraping.

Table of Contents

Title Page	1
Acknowledgment	2
Abstract.....	4
Table of Contents	5
List of Figures.....	7
List of Tables.....	8
Chapter 1: Introduction	9
1.1. Background and Significance of the Topic	9
1.2. Problem Statement.....	10
1.3. Research Objectives	11
1.4. Research Scope and Limitations.....	11
1.5. Research Methodology	12
1.6. Research Timeline	12
1.7. Structure of the Thesis	13
Chapter 2: System Analysis.....	14
2.1. Feasibility Analysis.....	14
2.2. Requirement Analysis.....	16
2.3. Technical Approach Analysis.....	24
2.4. Chapter Summary	26
Chapter 3: Design	27
3.1. Libraries Used.....	27
3.2. Architectural Structure.....	28
3.3. Functional Structure Design	29
3.4. Module Design	31
3.5. Data Collection Module.....	35
3.6. User Interface Design	36
3.7. Scoring Algorithms.....	40
3.8. Chapter Summary	43
Chapter 4: System Implementation.....	44
4.1. Streamlit Implementation	44

4.2.	User Interface Design Implementation	46
4.3.	Scraping Implementation.....	49
4.4.	Data Preprocessing and Scoring System Implementation	56
4.5.	Show Results and Export Implementation.....	60
4.6.	Chapter Summary	62
Chapter 5: Testing.....		63
5.1.	Module Testing	63
5.2.	Function Testing	69
5.3.	Possible Issue Testing	70
Chapter 6: Conclusion.....		73
References.....		74

List of Figures

Figure 1 Research Timeline	12
Figure 2 – 1 System Flow Chart	17
Figure 2 – 2 System Module Activity Diagram	18
Figure 2 – 3 User Case Diagram	19
Figure 3 – 1 Architectural Design	28
Figure 3 – 2 Functional Structure Hierarchy Diagram.....	30
Figure 3 – 3 Scraping Module Diagram	32
Figure 3 – 4 Data Preprocessing and Scoring System Module Diagram.....	33
Figure 3 – 5 Show Results and Export Module Diagram	34
Figure 3 – 6 Landing Page.....	37
Figure 3 – 7 Input Panel.....	37
Figure 3 – 8 Sorting Method Selection	37
Figure 3 – 9 Result Page.....	38
Figure 3 – 10 Product Information Card.....	39
Figure 4 – 1 Landing Page Implementation.....	47
Figure 4 – 2 Result Page Implementation.....	48
Figure 5 – 1 Scraping Result.....	63
Figure 5 – 2 Data Preprocessing Result.....	64
Figure 5 – 3 Export Raw Data	65
Figure 5 – 4 ‘Baju Erigo’ Search Result	65
Figure 5 – 5 Sorting Type Dropdown Menu	66
Figure 5 – 6 Changing Sort Method Result	66
Figure 5 – 7 Filtering Seller Dropdown Menu.....	67
Figure 5 – 8 Filtering Seller Result.....	67
Figure 5 – 9 Filtering Location Dropdown Menu.....	68
Figure 5 – 10 Filtering Location Result.....	68

List of Tables

Table 2 – 1 Search Item Specification Table.....	20
Table 2 – 2 View Result Specification Table	21
Table 2 – 3 Export Raw Data Specification Table	21
Table 2 – 4 Technical Approach Comparison Table.....	24
Table 3 – 1 Data Design Table	35
Table 3 – 2 Additional Score Data Table.....	36
Table 4 – 1 Streamlit Widgets Implemented in the Project Table	45
Table 5 – 1 Testing Conclusions Table.....	69
Table 5 – 2 Possible Issues and Bug Analysis Table.....	71