

## **LAPORAN PRAKTIKUM UAS KELOMPOK-3**

### **PEMROSESAN DATA**

- Eksplorasi Data Harga Tiket Pesawat melalui Web Scrapping -

Dosen Pengampu: Alfi Fadliana S.Si., M.Stat



Anggota Kelompok:

Arifah Shafyan (3322600003)

Ardiansyah Indra F. (3322600014)

Bayu Kurniawan (3322600019)

**PROGRAM STUDI D4 SAINS DATA TERAPAN**

**DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER**

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

**2023**

## DAFTAR ISI

DAFTAR ISI .....	1
DAFTAR GAMBAR.....	2
BAB I PENDAHULUAN .....	3
1.1 Latar Belakang .....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Pendekatan Konseptual Web Scraping .....	5
2.2 Aspek Legalitas Web scraping .....	6
2.3 Pemesanan tiket pesawat.....	6
2.4 Penelitian Terdahulu.....	7
BAB III METODE PENELITIAN .....	8
BAB IV PEMBAHASAN .....	9
BAB V KESIMPULAN .....	13
DAFTAR PUSTAKA .....	14
LAMPIRAN .....	16

## **DAFTAR GAMBAR**

Gambar i. Pendekatan Konseptual Pengembangan Web Scraping.....	5
Gambar ii. Harga Tiket Super Jet Air Berdasarkan Jam Yang Tersedia .....	9
Gambar iii. Harga Tiket Batik Air Berdasarkan Jam Yang Tersedia.....	9
Gambar iv. Harga Tiket Lion Air Berdasarkan Jam Yang Tersedia.....	10
Gambar v. Harga Tiket Citilink Berdasarkan Jam Yang Tersedia .....	10
Gambar vi. Harga Tiket Pelita Air Berdasarkan Jam Yang Tersedia .....	11
Gambar vii. Harga Tiket Garuda Indonesia Berdasarkan Jam Yang Tersedia .....	11

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dalam era digital yang semakin berkembang, akses mudah ke informasi merupakan hal yang penting dalam berbagai aspek kehidupan, termasuk dalam perencanaan perjalanan. Salah satu hal yang sangat dibutuhkan oleh para pelancong adalah informasi mengenai harga tiket pesawat. Dalam industri penerbangan, harga tiket pesawat dapat berubah secara dinamis dan sering kali dipengaruhi oleh berbagai faktor, seperti musim liburan, permintaan penumpang, dan kebijakan maskapai penerbangan.

Dalam rangka memperoleh informasi terkini mengenai harga tiket pesawat, banyak orang mengandalkan situs web seperti Tiket.com yang menyediakan fasilitas pencarian tiket pesawat online. Namun, mencatat harga tiket secara manual dari situs web seperti Tiket.com bisa menjadi tugas yang rumit dan memakan waktu.

Untuk mengatasi tantangan tersebut, teknik *web scraping* dapat menjadi solusi yang efisien. *Web scraping* adalah teknik pengumpulan data otomatis dari halaman web dengan menggunakan perangkat lunak atau script khusus. Dengan menerapkan teknik *web scraping* pada situs web Tiket.com, kita dapat mengambil data harga tiket pesawat secara otomatis, menyimpannya dalam bentuk dataset, dan kemudian menganalisisnya untuk mendapatkan wawasan yang berharga.

Laporan ini bertujuan untuk menjelaskan secara detail tentang penerapan teknik *web scraping* pada web Tiket.com untuk mengumpulkan dataset harga tiket pesawat Surabaya-Jakarta pada 31 Mei 2023. Laporan ini akan membahas langkah-langkah yang diperlukan dalam penerapan web scraping. Selain itu, Laporan ini juga akan membahas potensi aplikasi dari dataset yang dihasilkan, seperti analisis tren harga tiket pesawat, perbandingan harga antara maskapai penerbangan, dan penerapan metode analisis data lainnya. Laporan ini diharapkan dapat memberikan pemahaman yang lebih baik tentang penerapan teknik *web scraping* pada industri penerbangan dan memberikan manfaat bagi mereka yang tertarik dengan analisis data harga tiket pesawat.

Tantangan utama dalam penerapan teknik web scraping adalah menjaga etika dan kepatuhan hukum. Saat melakukan web scraping, penting untuk mematuhi aturan dan kebijakan situs web yang diakses, serta memperhatikan hak cipta dan *privasi data*. Laporan ini akan membahas tentang prinsip-prinsip etika dalam *web scraping*, serta memberikan saran praktis untuk memastikan penggunaan yang bertanggung jawab dan legal.

Selanjutnya, penting juga untuk mengatasi masalah validitas dan keandalan data yang diperoleh melalui web scraping. Dataset yang dihasilkan harus melalui proses pembersihan dan validasi agar dapat dipercaya dalam analisis dan pengambilan keputusan. Laporan ini akan

membahas strategi untuk mengatasi masalah ini, seperti penanganan data yang hilang atau tidak lengkap, penghapusan outlier, dan teknik validasi data.

Dengan menggabungkan langkah-langkah teknis, pertimbangan etika, validitas data, dan kebijakan penggunaan situs web, laporan ini akan memberikan panduan komprehensif tentang penerapan teknik web scraping pada situs web Tiket.com untuk mengumpulkan dataset harga tiket pesawat Surabaya-Jakarta pada 31 Mei 2023. Laporan ini diharapkan dapat menjadi sumber referensi yang berguna bagi para peneliti, analis, dan praktisi di bidang analisis data, serta menyumbangkan pemahaman yang lebih baik tentang potensi dan tantangan dalam penggunaan teknik web scraping pada industri penerbangan.

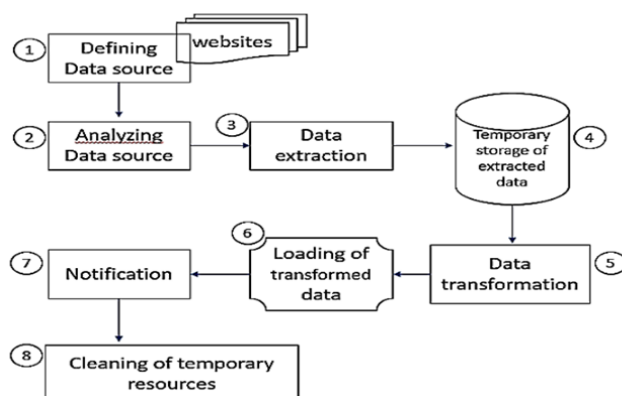
## BAB II TINJAUAN PUSTAKA

### 2.1 Pendekatan Konseptual Web Scraping

*Web scraping* adalah proses pengambilan sebuah dokumen semi dari halaman web, untuk diambil data tertentu saja dari halaman tersebut (Turland, 2010). *Web scraping* adalah teknik untuk mengekstraksi data dari internet dan disimpan ke *file* atau *database* untuk kebutuhan analisis data (Zhao, 2017). *Web scraping* sebagai teknik yang efisien dan kuat untuk mengumpulkan *big data* (Mooney et al. 2015; Bar-Ilan 2001 dalam Zhao, 2017). *Web scraping* dapat melakukan pencarian apa-apa yang tidak dapat dilakukan oleh mesin pencari tradisional seperti *Google Search* (Mitchell, 2015). Beberapa pengertian di atas menjelaskan bahwa *web scraping* adalah metode untuk mengekstraksi informasi dari situs web, sehingga menjadi data yang dapat dianalisis dan dimanfaatkan untuk berbagai tujuan.

Proses web scraping dari internet pada dasarnya terbagi menjadi dua langkah yang berurutan, yaitu menemukan web yang akan diekstrak datanya dan kemudian mengekstraksi data/informasi yang dibutuhkan dari web tersebut (Zhao, 2017). Beberapa *tools/software* yang dapat digunakan untuk melakukan *web scraping* adalah pemrograman python (Mitchell, 2015), PHP (Thurland, 2010), Bahasa R (Munzert, et.al, 2015), Microsoft Excel (Phillips, 2016), Botsol (Botsol.com, 2020), Outwit Hub, Spinn3r, Fminer, ParseHub, OctoParse, Import.io, VisualScaper, Dexi.io, 80legs dan Web Scraper (Milev, 2017). Dalam penelitian ini digunakan *Web Scraper* dan *Instant Data Scraper* sebagai ekstensi Google Chrome. Walaupun menurut Milev (2017) *Web Scraper* dan *Instant Data Scraper* ini termasuk aplikasi *web scraping* yang tradisional, namun aplikasi ini menawarkan kemudahan bagi penggunaanya, dan tanpa harus memahami bahasa pemrograman (coding).

Berbeda dengan sumber-sumber di atas, Milev (2017) menyebutkan *web scraping* hanyalah langkah awal dalam *data mining*, sementara *data mining* adalah bagian dari *business intelligence*(BI). Sistem BI adalah suatu sistem dinamis yang terintegrasi yang terus menerus berkembang di mana data dan fungsionalitas dikembangkan dalam beberapa iterasi berdasarkan umpan balik yang diperoleh dari penggunaanya. Dengan adanya system tersebut Milev (2017) mengusulkan pendekatan konseptual untuk pengembangan aplikasi *web scraping* dengan delapan langkah sebagai berikut:



**Gambar i.** Pendekatan Konseptual Pengembangan *Web Scraping*

- 1) Melakukan pendefinisian situs web atau menentukan *url* dari *website* yang akan diekstrak datanya;
- 2) Menganalisis sumber data yang didefinisikan untuk memastikan proses *web scraping* itu dapat dilakukan pada *web* tersebut dengan menyesuaikan bahasa yang digunakan dalam *web* tersebut atau menambahkan perintah tertentu dalam *script web scraping* untuk proses ekstraksinya.
- 3) Memulai proses ekstraksi data secara otomatis atau terjadwal;
- 4) Menyimpan data yang diekstraksi menjadi basis data temporary selama proses *web scraping*;
- 5) Transformasi hasil ekstraksi data ke dalam suatu bentuk yang bisa di-load ke dalam struktur data warehouse;
- 6) Load data yang telah ditransformasi ke dalam data warehouse;
- 7) Setelah data ada dalam *data warehouse*, dimungkinkan pencarian data berdasarkan daftar kata kunci yang diinginkan dan sistem akan memberikan notifikasi; dan
- 8) Pembersihan *temporary data*, sehingga hanya data yang telah ditransformasi yang di-load ke dalam *data warehouse* akan digunakan.

## 2.2 Aspek Legalitas Web scraping

Menurut beberapa pakar melakukan *web scraping* terhadap suatu web masih kontroversial. Menurut Fisher et al. (2010), Ketentuan Layanan (*term of service-ToS*) menciptakan hambatan spesifik untuk mengumpulkan informasi sebagai contoh ToS Facebook yang menyebutkan “*You will not collect users' information, or otherwise access Facebook, using automated means (such as harvesting bots, robots, spiders, or scrapers) without our permission.*”

Hirschey (2014) mengatakan *scraping* yang dapat merusak web dapat diadukan sebagai pelanggaran hak akses. O'Reilly (2006) menyebutkan *scraping* terkait dengan hak cipta (*intellectual property*). Krotov dan Silva (2018) menyebutkan bahwa aspek legalitas dan etika melakukan *web scraping* ini masih *grey area*.

Pada dasarnya teknik *web scraping* adalah proses *copy-paste* dari suatu web namun dilakukan secara otomatis dan terorganisir (Saurkar, et.all, 2018). Teknik *web scraper* mengambil data yang tersaji (*publish*) dalam suatu web, bukan merupakan teknik *hacking* yang mengambil data langsung dari server/database pemilik web (*backward database*). Dengan demikian, baik pemilik web maupun yang melakukan *scraping*, kedua belah pihak sebenarnya mendapat keuntungan, pihak pemilik webmendapat *traffics* yang dapat meningkatkan valuasinya sementara pihak yang melakukan *scraping* mendapat data yang dibutuhkan. Walaupun di sisi lain, penggunaan *web scraping* dapat mempengaruhi performa akses bagi pengguna lainnya.

## 2.3 Pemesanan tiket pesawat

Penerbangan indonesia saat ini telah mengalami kemajuan yang pesat khususnya dalam pemesanan tiket pesawat. Dengan menggunakan sistem informasi kita sangat mudah mendapat

informasi mengenai penerbangan kepada calon penumpang. Pemanfaatan teknologi juga mampu mengurangi jumlah SDM karena banyak pekerja rutinitas dapat digantikan oleh system aplikasi di komputer. Oleh karena itu ini menjadi tantangan bagi manajemen perusahaan penerbangan dalam persaingan yang sangat kompetitif antar maskapai sehingga maskapai dituntut untuk selalu melakukan langkah-langkah penghematan biaya pada perusahaan penerbangan.

Pada saat ini pesan tiket pesawat bisa dilakukan secara online. Beberapa platform sudah menyediakan layanan pemesanan tiket pesawat yang dapat diakses melalui smartphone. Dengan adanya cara pesan tiket pesawat secara online, calon penumpang tidak perlu repot datang ke kantor maskapai atau agen travel. Cukup memanfaatkan smartphone, pemesanan tiket pesawat akan selesai dalam hitungan menit. Jika beruntung, calon penumpang berkesempatan mendapat harga tiket pesawat murah atau tawaran promo menarik lainnya dari aplikasi atau marketplace tertentu.

## **2.4 Penelitian Terdahulu**

Josi, dkk (2014) melakukan penelitian implementasi *web scraping* pada sejumlah mesin pencari yang gratis seperti portal Garuda, Indonesian Scientific Journal Database (ISJD), dan Google Scholar, dengan hasil yang menyatakan bahwa dengan menerapkan teknik web scraping mampu mengekstrak informasi mengenai artikel jurnal ilmiah dari portal-portal tersebut sekaligus mampu menyimpan data hasil scraping pada database secara otomatis.

Sonya dan Prihandoko, (2016) melakukan penelitian pada tiga media online yaitu, Detikcom, Liputan6 dan VivaNews untuk mencari dan menganalisis informasi terkait bencana alam, dan menemukan bahwa Detikcom lebih banyak menghasilkan data yang relevan dibandingkan dengan Liputan6 dan VivaNews dan metode teknik *Breadth-First Search* sangat membantu dalam hal pencarian record dengan menelusuri URL utama hingga ke bagian terdalam link.

Fernando, et.al (2019) melakukan ekstraksi data produk di *marketplace* menggunakan web crawling dan *web scraping* dan memvisualisasikan hasilnya dengan aplikasi Power BI. Penelitian Fernando menunjukkan bahwa metode web crawling dan web scraping ini berguna untuk melakukan analisis data.



### BAB III METODE PENELITIAN

Harga Tiket Pesawat

Web yang bisa diakses:

<https://www.tiket.com/pesawat/search?d=SUBC&a=JKTC&dType=CITY&aType=CITY&date=2023-05-31&adult=1&child=0&infant=0&class=economy&flexiFare=true>

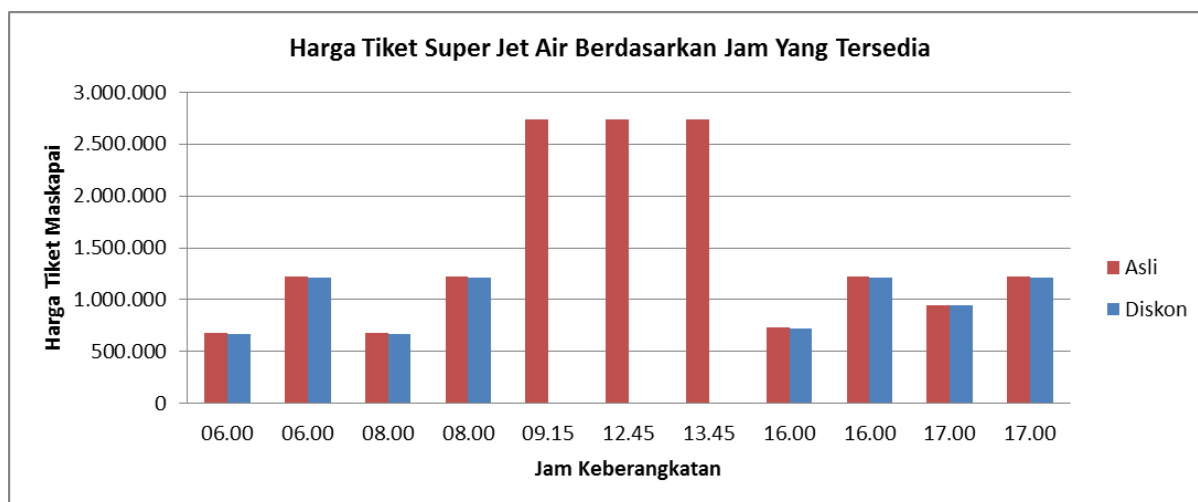
Kemudian dari halaman tersebut mencari seluruh nama maskapai yang tersedia untuk jadwal pemberangkatan tersebut, kemudian mencari waktu keberangkatan dan asal keberangkatan, dilanjutkan dengan mencari waktu tiba pada bandara tujuan dan juga bandara mana yang menjadi tujuan dari konsumen, dan beberapa variabel terakhir adalah metode perjalanan serta harga diskon dan juga harga asli dari setiap maskapai. Setiap variabel dicari berdasarkan filterisasi data pada Harga Tiket Maskapai pada tanggal 31 Mei 2023.

Setelah selesai dilakukan langkah-langkah coding, maka akan dihasilkan dataset pada file csv yang berisikan 7 variabel, yaitu Maskapai, Jam Takeoff, Bandara Takeoff, Metode Penerbangan, Jam Landing, Bandara Landing, Harga Diskon dan Harga Asli. Untuk melakukan analisis bisa cukup dengan menggunakan excel untuk menampilkan plot seperti grafik yang menjelaskan terkait harga tiket dari masing-masing 6 maskapai, dimana grafik ini bertujuan memberikan informasi mengenai harga tiket terbaik ataupun cocok pada keadaan konsumen di hari penerbangan dengan memberikan berbagai pilihan maskapai yang memberikan akomodir harga dan fasilitas yang dapat memberikan kenyamanan dan keamanan bagi konsumen.

Langkah awal dapat memahami data csv yang didapatkan, karena setiap variabel memiliki tipe data yang berbeda. Seringnya digunakan tipe data numerik untuk menciptakan visualisasi dengan grafik, oleh karena itu untuk menyampaikan informasi mengenai harga tiket yang mudah dipahami oleh pembaca akan dipilih grafik bar. Untuk menunjukkan hasilnya diperlukan memilih variabel dengan tipe numerik, hal ini ditunjukkan dengan pemilihan harga diskon dan harga asli yang menjadi sumbu-Y sedangkan pada sumbu-X digunakan nilai variabel jam pada setiap maskapai.

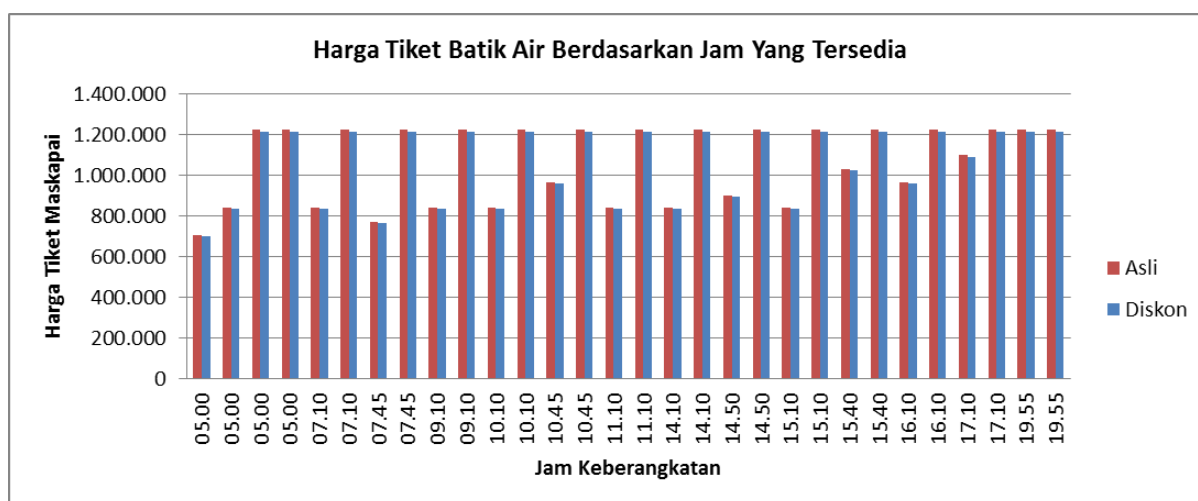
Langkah berikutnya adalah memberikan hasil visualisasi plot dengan masing-masing maskapai agar dapat digunakan sebagai perbandingan harga, dimana harga setiap maskapai akan berbeda-beda hal ini bisa dipengaruhi oleh pesawat tipe apa yang digunakan seperti Boeing atau Airbus. Ada juga *additional thing* layaknya tambahan kapasitas pada kabin, atau ada juga makanan gratis yang diberikan apabila memilih penerbangan dengan maskapai dan juga waktu tertentu.

## BAB IV PEMBAHASAN



**Gambar ii.** Harga Tiket Super Jet Air Berdasarkan Jam Yang Tersedia

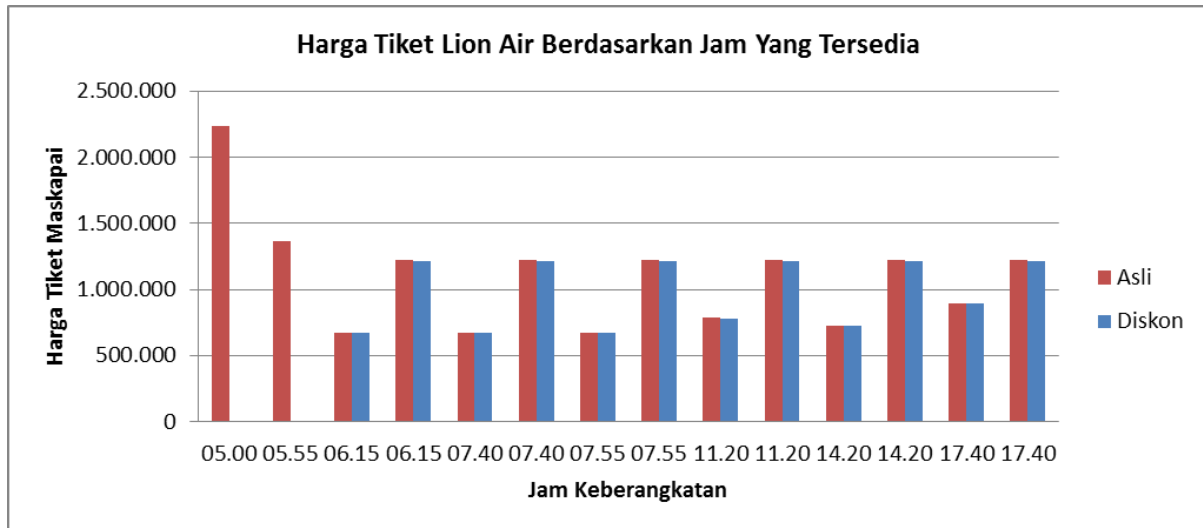
Pada plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Super Jet Air yang disesuaikan dengan jam keberangkatan. Dari harga diskon dan harga asli nampak tidak terlalu jauh berbeda harga yang diberikan. Ada juga grafik yang menjulang tinggi berbeda dengan yang lain serta tidak ada grafik yang menggambarkan diskon hal ini dipengaruhi beberapa hal seperti dari tipe pesawat serta keistimewaan yang dimiliki oleh penerbangan dari maskapai tersebut.



**Gambar iii.** Harga Tiket Batik Air Berdasarkan Jam Yang Tersedia

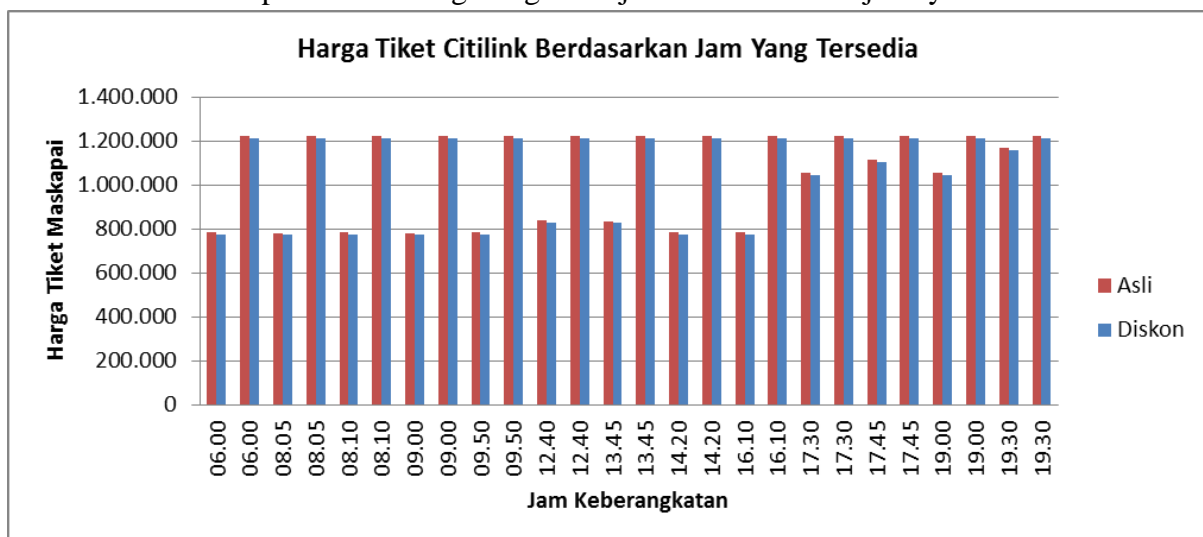
Plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Batik Air yang disesuaikan dengan jam keberangkatan. Dari harga diskon dan harga asli nampak tidak terlalu jauh berbeda harga yang diberikan. Ada juga grafik yang menjulang tinggi daripada beberapa grafik diantaranya hal ini dipengaruhi beberapa hal seperti dari tipe pesawat serta keistimewaan yang dimiliki oleh penerbangan dari maskapai

tersebut. Serta apabila diamati waktu atau jam penerbangan yang ditampilkan harga yang dimiliki oleh maskapai tersebut tergolong tidak jauh berbeda antar jamnya.



**Gambar iv.** Harga Tiket Lion Air Berdasarkan Jam Yang Tersedia

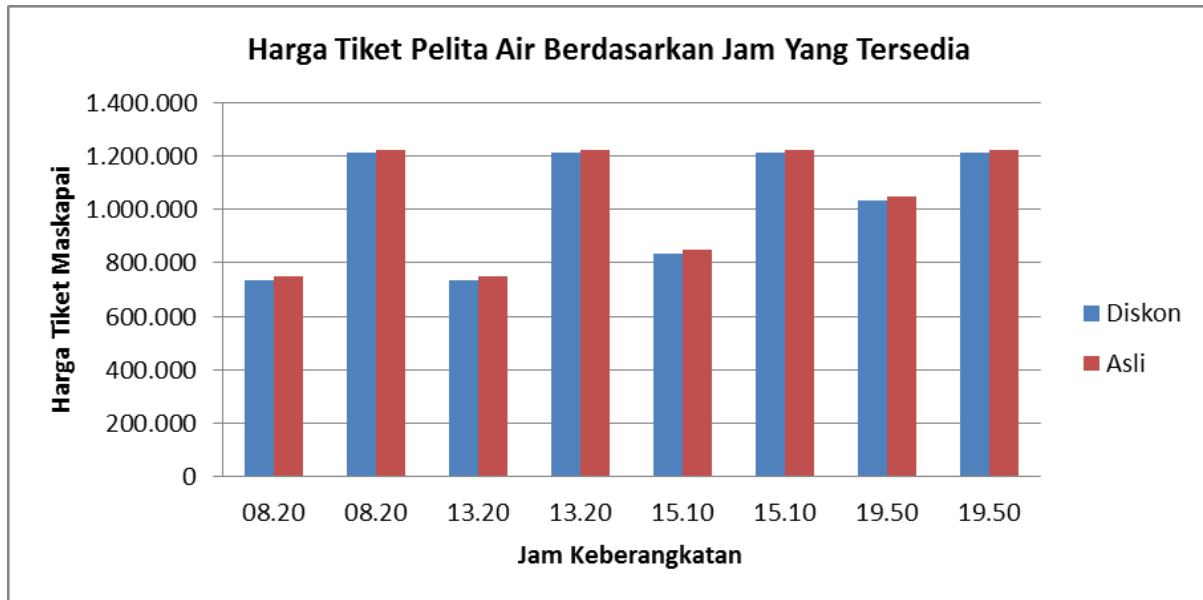
Plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Lion Air yang disesuaikan dengan jam keberangkatan. Dari harga diskon dan harga asli nampak tidak terlalu jauh berbeda harga yang diberikan. Ada juga grafik yang menjulang tinggi daripada beberapa grafik diantaranya hal ini dipengaruhi beberapa hal seperti dari tipe pesawat serta keistimewaan yang dimiliki oleh penerbangan dari maskapai tersebut. Serta apabila diamati waktu atau jam penerbangan yang ditampilkan harga yang dimiliki oleh maskapai tersebut tergolong tidak jauh berbeda antar jamnya.



**Gambar v.** Harga Tiket Citilink Berdasarkan Jam Yang Tersedia

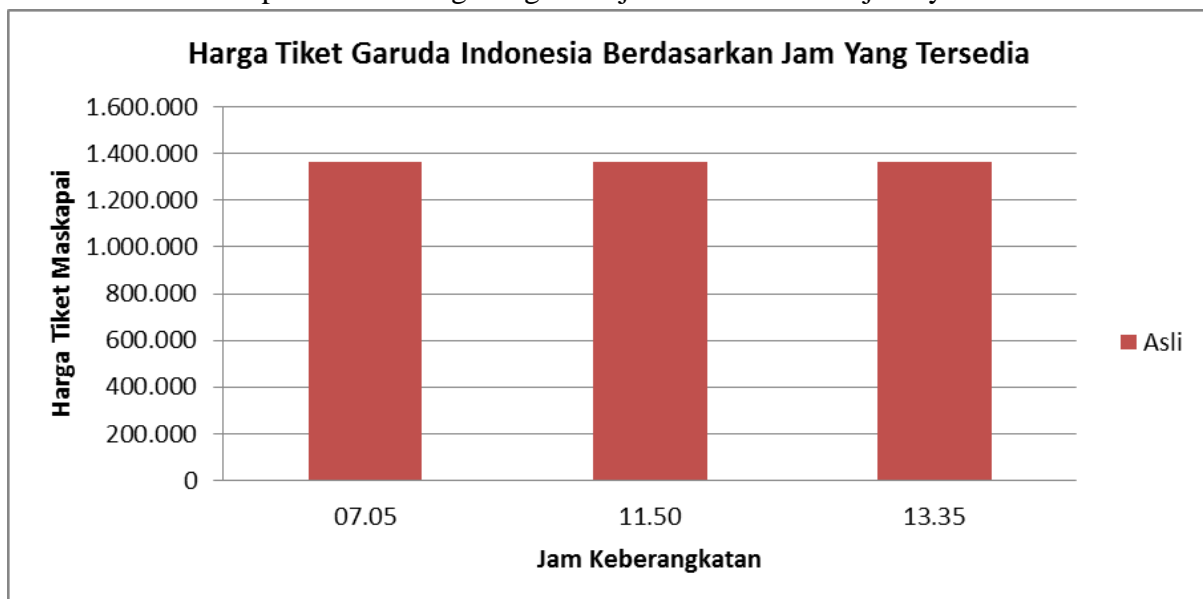
Plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Citilink yang disesuaikan dengan jam keberangkatan. Dari harga diskon dan harga asli nampak tidak terlalu jauh berbeda harga yang diberikan. Ada juga grafik yang menjulang tinggi daripada beberapa grafik diantaranya hal ini dipengaruhi beberapa hal seperti dari tipe pesawat serta keistimewaan yang dimiliki oleh penerbangan dari maskapai

tersebut. Serta apabila diamati waktu atau jam penerbangan yang ditampilkan harga yang dimiliki oleh maskapai tersebut tergolong tidak jauh berbeda antar jamnya.



**Gambar vi.** Harga Tiket Pelita Air Berdasarkan Jam Yang Tersedia

Plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Pelita Air yang disesuaikan dengan jam keberangkatan. Dari harga diskon dan harga asli nampak tidak terlalu jauh berbeda harga yang diberikan. Ada juga grafik yang menjulang tinggi daripada beberapa grafik diantaranya hal ini dipengaruhi beberapa hal seperti dari tipe pesawat serta keistimewaan yang dimiliki oleh penerbangan dari maskapai tersebut. Serta apabila diamati waktu atau jam penerbangan yang ditampilkan harga yang dimiliki oleh maskapai tersebut tergolong tidak jauh berbeda antar jamnya.



**Gambar vii.** Harga Tiket Garuda Indonesia Berdasarkan Jam Yang Tersedia

Plot diatas menampilkan grafik dari harga tiket penerbangan dari Surabaya menuju ke Jakarta dengan maskapai Garuda Indonesia yang disesuaikan dengan jam keberangkatan.

Harga asli nampak tidak terlalu jauh berbeda harga yang diberikan pada setiap jam yang ada tersedia pada maskapai tersebut. Apabila diamati dengan sekilas, nampak harga dari Garuda Indonesia hanya memiliki 1 kelas penerbangan tidak cukup banyak variasi yang diberikan pada tiketnya. Hal tersebut bisa dipengaruhi karena *demand* dari warga domestik tidak begitu banyak.

## **BAB V KESIMPULAN**

Dalam mencari tiket pesawat termurah dari Surabaya ke Jakarta, terdapat beberapa faktor yang perlu dipertimbangkan. Pertama, pemilihan waktu penerbangan yang tepat menjadi penting. Menghindari periode liburan atau akhir pekan dapat membantu menemukan tiket dengan harga lebih terjangkau. Hal ini dapat menjadi pertimbangan bagi para pelajar layaknya mahasiswa untuk memesan tiket pesawat pada jauh-jauh sebelum mendekati waktu dari liburan akan datang. Kedua, mencari promosi, diskon khusus, dan memanfaatkan kode promo serta program loyalitas dapat membantu mengurangi biaya perjalanan. Fleksibilitas dalam jadwal penerbangan juga bisa menjadi kunci. Memilih penerbangan pada waktu yang tidak populer dapat menghemat uang. Selanjutnya, membandingkan harga tiket dari berbagai maskapai dan agen perjalanan adalah langkah yang penting. Situs perbandingan harga atau aplikasi pemesanan tiket dapat membantu menemukan penawaran terbaik. Terakhir, perhatikan juga tambahan biaya seperti bagasi terdaftar, makanan di pesawat, atau pemilihan kursi. Memperhitungkan semua biaya tersebut dalam memilih tiket akan membantu mendapatkan harga tiket pesawat yang lebih terjangkau secara keseluruhan. Dengan mempertimbangkan faktor-faktor ini, penumpang dapat meningkatkan peluang mereka untuk mendapatkan tiket pesawat termurah dan mengoptimalkan pengeluaran mereka dalam perjalanan dari Surabaya ke Jakarta.

Sehingga kesimpulan dari hasil scrapping web tiket.com dihasilkan dataset yang berisikan perbandingan harga tiket beberapa maskapai dengan masing-masing jam penerbangan dan keunggulan dari masing-masing maskapai. Kami coba golongan beberapa penerbangan dengan 2 kategori jam yang berbeda, masing-masing digolongkan pada pukul diatas jam 12.00 WIB dan dibawah dari jam 12.00 WIB. Dari pada jam sebelum 12.00 WIB didapatkan bahwa terdapat tiket penerbangan dari maskapai Super Air Jet pada jam keberangkatan 08.00 hingga jam kedatangan 09.25 dengan harga Rp. 670.771. Jika dilihat dari penerbangan diatas pukul 12.00 WIB terdapat Super Jet Air dimana jam keberangkatan 16.00 WIB dengan jam kedatangan pukul 17.30 dimiliki tiket termurah pada kategori pukul tersebut seharga Rp. 725.966. Dengan hasil penelitian ini diharapkan dapat memberikan informasi terkait tiket penerbangan dengan beberapa maskapai yang bisa dijadikan variasi pilihan untuk menentukan tiket mana yang akan dibeli dan kami juga memberikan beberapa saran sebagai opsi dari pembelian tiket termurah dari 2 kategori jam yang ditampilkan.

## DAFTAR PUSTAKA

- Fernando, Edward Hanafi, et.al, 2019, “Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling”, Jurnal Teknik Informatika dan Sistem Informasi, Volume 5 Nomor 3 Desember 2019.
- Fisher, D., McDonald, D. W., Brooks, A. L., & Churchill, E. F. , 2010, “Terms of service, Ethics, and Bias: Tapping the social web for CSCW research”, CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.
- Hirschey, J. K., 2014. “Symbiotic Relationships: Pragmatic Acceptance of Data Scraping”, Berkeley Technology Law Journal, Vol. 29, 2014.
- Josi, A., Abdillah L.A., Suryayusra, 2014, Penerapan Teknik Web scraping pada Mesin Pencari Artikel Ilmiah, dalam laman [https://www.researchgate.net/publication/267214300\\_Penerapan\\_teknik\\_web\\_scraping\\_pada\\_mesin\\_pencari\\_artikel\\_ilmiah](https://www.researchgate.net/publication/267214300_Penerapan_teknik_web_scraping_pada_mesin_pencari_artikel_ilmiah) diakses tanggal 2 Juni 2023.
- Krotov, Vlad dan Leiser Silva, 2018, “Legality And Ethics Of Web Scraping”, Emergent Research Forum (ERF), Twenty-fourth Americas Conference on Information Systems, New Orleans,
- Milev, Plamen, 2017, “Conceptual Approach for Development of Web scraping Application for Tracking Information”, Economic Alternatives, 2017, Issue 3, pp. 475-485.
- Mitchell, Ryan, 2015, Web scraping with Python Collecting Data from the Modern Web, O’Reilly, USA.
- Munzert, Simon, Christian Rubba, Peter Meißner, Dominic Nyhuis, 2015, Automated Data Collection with R: A Practical Guide to Web scraping and Text Mining, Wiley.
- O’Reilly, S., (2006, “Nominative Fair Use and Internet Aggregators: Copyright and Trademark Challenges Posed by Bots, Web Crawlers and Screen-Scraping Technologies”, Loyola Consumer Law Review Volume 19 Issue 3.
- Phillips, David M. W, 2016, Web scraping with Excel : How to Use VBA to Write Simple and Effective Web Scrapes, CreateSpace Independent Publishing Platform.
- Saurkar, Anand V., Kedar G. Pathare, Shweta A. Gode, 2018, “An Overview On Web scraping Techniques And Tools”, International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCCE), Volume: 4 Issue: 4, April 2018.
- Sonya, Izatul Putri dan Prihandoko, 2016, “Analisis Web scraping Untuk Data Bencana Alam Dengan Menggunakan Teknik Breadth-First Search Terhadap 3 Media Online”, Jurnal Informatika dan Komputer Volume 21 No. 3, Desember 2016.

Turland, M., 2010, *php| architect's Guide to Web scraping*, Marco Tabini & Associates, Inc, USA.

Tsani, Irlita. M. F, 2017, *Penerapan Teknik Web Scraping Pada Website Tiket Pesawat Sebagai Pembanding Harga Tiket Pesawat Berbasis Web Responsive*, UMM, Malang.

Zhao, Bo, 2017, *Web scraping*, Springer International Publishing AG (outside the USA) 2017.



## LAMPIRAN

```
import time
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By

url =
"https://www.tiket.com/pesawat/search?d=SUBC&a=JKTC&dType=CITY&aType=CI
TY&date=2023-05-
31&adult=1&child=0&infant=0&class=economy&flexiFare=true"
driver = webdriver.Chrome()
driver.get(url)

data = []

for j in range(45):
    time.sleep(1)
    driver.execute_script("window.scrollTo(0, 250)")
    time.sleep(1)

soup = BeautifulSoup(driver.page_source, "html.parser")

#Discount Ticket
for item in soup.findAll("div", class_ = "harga_coret_regular"):
    maskapai = item.find("span", class_ = "text-marketing-
airline").text
    print(maskapai)
    print("=")
    jam_takeoff = item.find("div", class_ = "text-time").text
    print(jam_takeoff)
    print("==")
    bandara_takeoff = item.find("div", class_ = "text-code").text
    print(bandara_takeoff)
    print("===")

    for item_2 in item.findAll("div", class_ = "list-
horizontal__middle"):
        for i,item__1 in enumerate(item_2.findAll("div", class_ =
"row")):
            if i == 0:
```

```

        jam_penerbangan = item__1.find("div", class_ = "text-
total-time").text
        print(jam_penerbangan)
        print("====")
        break

    for j,item__2 in enumerate(item_2.findAll("div", class_ =
"row")):
        if j == 2:
            metode_penerbangan = item__2.find("div", class_ =
"text-total-time").text
            print(metode_penerbangan)
            print("=====")
            break

    for k, item_3 in enumerate(item.findAll("div", class_ = "list-
horizontal__middle")[4]):
        if k == 0:
            jam_landing = item_3.text
            print(jam_landing)
            print("=====")
        if k == 1:
            bandara_landing = item_3.text
            print(bandara_landing)
            print("=====")
            break

    for item_4 in item.findAll("div", class_ = "priceV2"):
        for l,item__3 in enumerate(item_4.findAll("div", class_ =
"text-price")):
            if l == 0:
                harga_diskon = item__3.find("div", class_ = "text-
price-main").text
                print(harga_diskon)
                print("=====")
            if l == 1:
                harga_asli = item__3.find("div", class_ = "text-price-
main").text
                print(harga_asli)
                print("=====")
                break

    data.append(

```

```

        (maskapai, jam_takeoff, bandara_takeoff,
metode_penerbangan, jam_landing, bandara_landing, harga_diskon,
harga_asli)
    )

#Not Discount Ticket
for item2 in soup.findAll("div", class_ = "wrapper-flight-list"):
    maskapai = item2.find("span", class_ = "text-marketing-
airline").text
    print(maskapai)
    print("-")
    jam_takeoff = item2.find("div", class_ = "text-time").text
    print(jam_takeoff)
    print("--")
    bandara_takeoff = item2.find("div", class_ = "text-code").text
    print(bandara_takeoff)
    print("---")

    for item_5 in item2.findAll("div", class_ = "list-
horizontal__middle"):
        for i,item__4 in enumerate(item_5.findAll("div", class_ =
"row")):
            if i == 0:
                jam_penerbangan = item__4.find("div", class_ = "text-
total-time").text
                print(jam_penerbangan)
                print("----")
                break
            for j,item__5 in enumerate(item_5.findAll("div", class_ =
"row")):
                if j == 2:
                    metode_penerbangan = item__5.find("div", class_ =
"text-total-time").text
                    print(metode_penerbangan)
                    print("-----")
                    break

            for k, item_6 in enumerate(item2.findAll("div", class_ = "list-
horizontal__middle")[4]):
                if k == 0:
                    jam_landing = item_6.text
                    print(jam_landing)
                    print("-----")

```

```

        if k == 1:
            bandara_landing = item_6.text
            print(bandara_landing)
            print("-----")
            break

    for item_7 in item2.findAll("div", class_ = "priceV2"):
        for item__6 in item_7.findAll("div", class_ = "text-price-
default"):
            harga_diskon = " "
            print(harga_diskon)
            print("-----")
            harga_asli = item__6.find("div", class_ = "text-price-main-
default").text
            print(harga_asli)
            print("-----")

    data.append(
        (maskapai, jam_takeoff, bandara_takeoff,
metode_penerbangan, jam_landing, bandara_landing, harga_diskon,
harga_asli)
    )

df = pd.DataFrame(data, columns = ["Maskapai", "Jam Takeoff", "Bandara
Takeoff", "Metode Penerbangan", "Jam Landing",
"Bandara Landing", "Harga Diskon", "Harga Asli"])
print(df)

df.to_csv("tiketcom.csv", index = False)
print("Data Telah Tersimpan")
driver.close()

```