



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - IS184853

**ANALISIS PEMANFAATAN PLATFORM DARING
UNTUK MENGGAMBARAKAN KETERSEDIAAN
INFORMASI BISNIS PADA UMKM BIDANG MAKANAN
DAN MINUMAN DI SURABAYA**

***ANALYSIS OF UTILIZATION OF ONLINE PLATFORM
TO IDENTIFY THE AVAILABILITY OF BUSINESS
INFORMATION IN THE FIELD OF FOOD AND
BEVERAGE IN SURABAYA***

Ubai Yusuf Siraj Samudera
0521 16 4000 0023

Dosen Pembimbing
Rully Agus Hendrawan, S.Kom., M.Eng.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020

TUGAS AKHIR - IS184853

**ANALISIS PEMANFAATAN PLATFORM DARING
UNTUK MENGGAMBARKAN KETERSEDIAAN
INFORMASI BISNIS PADA UMKM BIDANG
MAKANAN DAN MINUMAN DI SURABAYA**

**Ubai Yusuf Siraj Samudera
0521 16 4000 0023**

**Dosen Pembimbing
Rully Agus Hendrawan, S.Kom., M.Eng.**

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020**

UNDERGRADUATE THESIS - IS184853

**ANALYSIS OF UTILIZATION OF ONLINE
PLATFORM TO IDENTIFY THE
AVAILABILITY OF BUSINESS INFORMATION
IN THE FIELD OF FOOD AND BEVERAGE IN
SURABAYA**

Ubai Yusuf Siraj Samudera
0521 16 4000 0023

Supervisor
Rully Agus Hendrawan, S.Kom., M.Eng.

INFORMATION SYSTEM DEPARTMENT
Faculty of Intelligent Electrical and Informatics Technology
Sepuluh Nopember Institute of Technology
Surabaya 2020

Dr. Mudjahidin, ST., MT.
NIP. 197010102003121001



LEMBAR PERSETUJUAN

ANALISIS PEMANFAATAN PLATFORM DARING UNTUK MENGGAMBARAKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM BIDANG MAKANAN DAN MINUMAN DI SURABAYA

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh:

Ubai Yusuf Siraj Samudera

NRP. 0521 16 4000 0023

Disetujui Tim Penguji : Tanggal Ujian: 03 Juni 2020

Periode Wisuda: 2020

Rully Agus Hendrawan, S.Kom., M.Eng. (Pembimbing I)

Mahendrawati ER, S.T., M.Sc., Ph.D (Penguji I)

Erma Suryani, S.T., M.T., Ph.D (Penguji II)

ANALISIS PEMANFAATAN PLATFORM DARING UNTUK MENGGAMBARAKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM BIDANG MAKANAN DAN MINUMAN DI SURABAYA

Nama Mahasiswa : Ubai Yusuf Siraj Samudera
NRP : 0521 16 4000 0023
Departemen : Sistem Informasi FTEIC-ITS
**Pembimbing I : Rully Agus Hendrawan, S.Kom.,
M.Eng.**

ABSTRAK

UMKM sebagai salah satu penyumbang besar terhadap ekonomi pembangunan nasional. Lebih spesifik, bidang makanan dan minuman menunjukkan tren yang meningkat terhadap kontribusi perekonomian nasional. Dalam proses bisnisnya, UMKM memiliki opsi yang dapat dilakukan dalam melakukan penjualan salah satunya dengan memanfaatkan platform daring. Platform daring memiliki fitur-fitur untuk memperlihatkan kondisi bisnis melalui data. Sehingga, platform daring dapat menjadi opsi sebagai media yang dapat meningkatkan potensi bisnis UMKM apabila dilakukan analisis terhadap data yang tersedia. Sayangnya, data yang tersedia terkait dengan kondisi UMKM masih terbatas dan terpisah pada platform-platform daring tertentu.

Pada penelitian ini dilakukan pengumpulan data dan analisis kondisi UMKM dari berbagai platform daring. Pengumpulan dan analisis data dilakukan sebagai usaha untuk mengidentifikasi kondisi pemanfaatan platform daring pada UMKM makanan dan minuman. Pengumpulan dan analisis data memiliki beberapa tujuan yang salah satunya diwujudkan dalam penelitian ini, yaitu analisis data UMKM untuk memberikan gambaran ketersediaan informasi bisnis dan kondisi UMKM, serta memberikan informasi kepada pengguna dalam menjangkau UMKM.

Analisis dilakukan menggunakan metode analisis deskriptif menggunakan data yang dikumpulkan melalui web scraping. Pengumpulan data diotomasi menggunakan kode dengan basis bahasa Python. Hasil data yang dikumpulkan dari berbagai platform kemudian diintegrasikan dan dilakukan pengujian dengan mengukur tingkat kebenaran data dari hasil integrasi. Hasil dari penelitian ini berupa kode pengambilan data pada website, kumpulan data UMKM dari berbagai platform daring, serta halaman visualisasi informasi. Sehingga dapat dilakukan analisis terhadap UMKM makanan dan minuman dari berbagai platform daring yang digunakan.

Kata Kunci: *UMKM, Platform Daring, Pengumpulan Data, Web Scraping, Integrasi, Visualisasi Informasi, Analisis Deskriptif*

**ANALYSIS OF UTILIZATION OF ONLINE
PLATFORM TO IDENTIFY THE AVAILABILITY OF
BUSINESS INFORMATION IN THE FIELD OF FOOD
AND BEVERAGE IN SURABAYA**

Name : Ubai Yusuf Siraj Samudera
NRP : 0521 16 4000 0023
Department : Information System FTEIC-ITS
Supervisor I : Rully Agus Hendrawan, S.Kom., M.Eng.

ABSTRACT

MSME as one of the major contributors to the national development economy. More specifically, food and beverage fields show an increased trend towards national economic contributions. In the business process, SMES have the option that can be done in selling one of them by utilizing the online platform. The online Platform has features to showcase business conditions through data. Thus, the online platform can be an option as a medium that can increase the potential of MSME business when analysis is conducted on available data. Unfortunately, the available data related to MSME conditions are still limited and separate to certain online platforms.

In this research, data collection and analysis of MSME condition from various online platforms. Data collection and analysis are conducted as an effort to identify the conditions of utilization of online platforms on MSME food and beverage. Data collection and analysis has a number of objectives, one of which is realized in this research, which is the MSME data analysis to provide an overview or identification of conditions and provide information to users in reaching MSME.

Analysis is performed using a descriptive analysis method using the data collected through web scraping. Data collection is automated using code with a Python language base. The results of data collected from various platforms are then integrated and

tested by measuring the level of data veracity of the integration results. The results of this study form the data retrieval code on the website, the MSMES data collection from various online platforms, as well as information visualization pages. So, it can be done analysis of SMES food and beverage from various online platforms used.

Keywords: *MSME, Online Platform, Data collection, Web Scraping, Integration, Information Visualization, Descriptive Analysis*

LEMBAR PERNYATAAN BEBAS PLAGIARISME

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertandatangan di bawah ini:

Nama : Ubai Yusuf Siraj Samudera
NRP : 05211640000023
Tempat/Tanggal lahir : Sidoarjo / 23 Maret 1998
Fakultas/Departemen : Fakultas Teknologi Elektro dan Informatika Cerdas /
Sistem Informasi
Nomor Telp/Hp/email : 082230707825 / ubaiyusuf@gmail.com

Dengan ini menyatakan dengan sesungguhnya bahwa penelitian/makalah/tugas akhir saya yang berjudul

**“ANALISIS PEMANFAATAN PLATFORM DARING UNTUK
MENGGAMBARAKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM
BIDANG MAKANAN DAN MINUMAN DI SURABAYA”**

Bebas Dari Plagiarisme Dan Bukan Hasil Karya Orang Lain.

Apabila di kemudian hari ditemukan seluruh atau sebagian penelitian/makalah/tugas akhir tersebut terdapat indikasi plagiarisme, maka saya bersedia menerima sanksi sesuai peraturan dan ketentuan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 07 Agustus 2020



Ubai Yusuf Siraj Samudera
NRP. 05211640000023

Halaman ini sengaja dikosongkan.

KATA PENGANTAR

Dengan mengucapkan rasa syukur kepada Tuhan Yang Maha Pengasih dan Maha Penyayang atas izin-Nya penulis dapat menyelesaikan buku yang sederhana ini dengan judul

**“ANALISIS PEMANFAATAN PLATFORM DARING
UNTUK MENGGAMBARAKAN KETERSEDIAAN
INFORMASI BISNIS PADA UMKM BIDANG
MAKANAN DAN MINUMAN DI SURABAYA”.**

Dalam penyelesaian Tugas Akhir ini, penulis diiringi oleh pihak-pihak yang selalu memberi dukungan, saran, dan doa sehingga penelitian berlangsung dengan lancar. Secara khusus penulis mengucapkan terima kasih dari lubuk hati terdalam kepada:

1. Tuhan, yang selalu menemani dan membimbing penulis dalam segala aspek kehidupan.
2. Orang tua penulis, yang tiada hentinya mendoakan dan memberikan dukungan kepada penulis.
3. Dr. Mudjahidin, S.T., M.T. selaku Ketua Departemen Sistem Informasi ITS Surabaya.
4. Rully Agus Hendrawan, S.Kom., M.Eng. selaku dosen pembimbing yang telah mencurahkan segenap tenaga, waktu dan pikiran dalam penelitian ini, serta memberikan motivasi yang membangun.
5. Ibu Erma Suryani, S.T., M.T., Ph.D dan Ibu Mahendrawathi ER. S.T., M.Sc., Ph.D. selaku dosen penguji yang telah memberikan kritik dan saran yang membuat kualitas penelitian ini lebih baik lagi.
6. Segenap dosen dan karyawan Departemen Sistem Informasi.
7. Zainuddin, Dimas, Ferdian, Saifuddin, Fauzi, Dzulfikar, dan Adam selaku sahabat yang senantiasa membantu

penulis menyelesaikan Tugas Akhir dalam bentuk diskusi, dan rekreasi.

8. Pihak lainnya yang berkontribusi dalam tugas akhir yang belum dapat penulis sebutkan satu per satu.

Penyusunan tugas akhir ini masih jauh dari kata sempurna, untuk itu penulis menerima segala kritik dan saran yang membangun sebagai upaya menjadi lebih baik lagi ke depannya. Semoga buku tugas akhir ini dapat memberikan manfaat untuk pembaca.

Surabaya, 07 Agustus 2020

Ubai Yusuf Siraj Samudera

DAFTAR ISI

LEMBAR PENGESAHAN.....	iii
LEMBAR PERSETUJUAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vii
LEMBAR PERNYATAAN BEBAS PLAGIARISME	ix
KATA PENGANTAR	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xvii
DAFTAR TABEL.....	xix
1 BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Permasalahan.....	3
1.4 Tujuan	4
1.5 Metode	4
1.6 Manfaat	4
1.7 Target Luaran.....	5
1.8 Relevansi.....	5
2 BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Sebelumnya.....	7
2.2 Usaha Mikro Kecil Menengah (UMKM).....	9
2.3 <i>Data Collection</i>	10
2.4 <i>Web Scraping</i>	10
2.5 <i>Extraction, Transformation, Loading (ETL)</i>	11
2.6 <i>Entity Resolution</i>	11
3 BAB III METODOLOGI.....	13
3.1 Diagram Metodologi.....	13
3.2 Uraian Metodologi.....	18
3.2.1 Studi Literatur	18
3.2.2 <i>Data Scraping</i>	18
3.2.3 <i>Pre-processing</i>	20
3.2.4 <i>Data Loading</i>	21
3.2.5 Pembuatan Matriks Visualisasi.....	21
3.2.6 Visualisasi dan Analisis	21
3.2.7 Pelaporan.....	22
3.3 Rangkuman Metodologi.....	22

4 BAB IV PERANCANGAN	25
4.1 Rancangan <i>Data Scraper</i>	25
4.1.1 Tokopedia.....	26
4.1.2 Shopee	35
4.1.3 Google My Business	43
4.1.4 Instagram.....	48
4.2 Rancangan <i>Data Integration</i>	50
4.3 Perhitungan Kualitas Data Integrasi.....	52
4.3.1 Menentukan <i>Confidence Level</i>	52
4.3.2 Perhitungan Jumlah <i>Sample</i>	52
4.3.3 Pengambilan <i>Sample</i>	53
4.3.4 Validasi data.....	53
4.3.5 Perhitungan Nilai Presisi.....	53
4.4 Rancangan Visualisasi dan Analisis.....	53
4.4.1 Matriks	54
4.4.2 Visualisasi Data.....	56
5 BAB V IMPLEMENTASI	59
5.1 <i>Scraping Data</i>	59
5.1.1 Tokopedia.....	60
5.1.2 Shopee	68
5.1.3 Google My Business	74
5.1.4 Instagram.....	78
5.2 <i>Preprocess Data</i>	80
5.2.1 Data UMKM Tokopedia	80
5.2.2 Data UMKM Shopee.....	83
5.2.3 Data UMKM Google My Business	84
5.3 Integrasi Data	86
5.4 Perhitungan Nilai Tingkat Kebenaran.....	88
5.5 Visualisasi dan Analisis	89
5.5.1 <i>Card</i>	89
5.5.2 <i>Gauge</i>	89
5.5.3 <i>Scatter Plot</i>	90
5.5.4 <i>Bar Chart</i>	90
5.5.5 <i>Map</i>	91
5.5.6 <i>Donut Chart</i>	92
5.5.7 Wordcloud.....	92
6 BAB VI HASIL DAN PEMBAHASAN.....	95

6.1	<i>Data</i>	95
6.1.1	Tokopedia	95
6.1.2	Shopee.....	104
6.1.3	Google My Business	112
6.2	Integrasi Data	113
6.2.1	Tokopedia, dan Shopee	114
6.2.2	Google My Business, Tokopedia, dan Shopee..	114
6.3	Visualisasi dan Analisis	114
7	BAB VII PENUTUP	119
7.1	Kesimpulan	119
7.2	Saran	122
	DAFTAR PUSTAKA	125
	LAMPIRAN A . KODE PROGRAM	129
	LAMPIRAN B . DATA HASIL <i>SCRAPING</i> , <i>PREPROCESS</i> , DAN INTEGRASI.....	134
	LAMPIRAN C . <i>TABEL CHI SQUARE</i>	139
	BIODATA PENULIS	141

Halaman ini sengaja dikosongkan.

DAFTAR GAMBAR

Gambar 1.1 Kerangka Kerja Riset Laboratorium Sistem <i>Enterprise</i>	6
Gambar 3.1 Metodologi Penelitian Bagian 1.	13
Gambar 3.2 Metodologi Penelitian Bagian 2.	14
Gambar 3.3 Metodologi Penelitian Bagian 3.	15
Gambar 3.4 Metodologi Penelitian Bagian 4.	16
Gambar 3.5 Metodologi Penelitian Bagian 5.	17
Gambar 4.1 <i>Work Flow Data Scraper</i>	25
Gambar 5.1 Ilustrasi Hasil Pencarian Awal di google.com....	76
Gambar 5.2 Ilustrasi Halaman Pencarian di Google My Business.....	77
Gambar 5.3 Ilustrasi Detil Informasi UMKM Google My Business.....	77
Gambar 5.4 Implementasi Visualisasi <i>Card</i>	89
Gambar 5.5 Implementasi Visualisasi <i>Gauge</i>	90
Gambar 5.6 Implementasi Visualisasi <i>Scatter Plot</i>	90
Gambar 5.7 Implementasi Visualisasi <i>Bar Chart</i>	91
Gambar 5.8 Implementasi Visualisasi <i>Map</i>	91
Gambar 5.9 Implementasi Visualisasi <i>Donut Chart</i>	92
Gambar 5.10 Implementasi Visualisasi Wordcloud.....	93
Gambar 6.1 Pratinjau Data Kategori Makanan dan Minuman Tokopedia.....	96
Gambar 6.2 Pratinjau Data Produk Kategori Makanan dan Minuman Tokopedia.	96
Gambar 6.3 <i>Wordcloud</i> Nama Produk Makanan dan Minuman Tokopedia.....	97
Gambar 6.4 Diagram Persentase Perbandingan <i>URL</i> Promo dan <i>URL</i> Normal.	98
Gambar 6.5 Pratinjau Data <i>URL</i> UMKM Makanan Minuman Tokopedia.....	99
Gambar 6.6 Diagram Persentase Perbandingan UMKM Promo dan UMKM Normal.	100
Gambar 6.7 Pratinjau Hasil <i>Scraping</i> Data Informasi UMKM Makanan dan Minuman Tokopedia.....	100
Gambar 6.8 Jumlah UMKM Berdasarkan Tipe <i>Seller</i>	101

Gambar 6.9 Persentase UMKM yang Telah Memiliki <i>Rating</i> dengan yang Belum Memiliki <i>Rating</i>	102
Gambar 6.10 Persentase Jumlah UMKM Tokopedia Berdasarkan Kecepatan Proses Transaksi.	103
Gambar 6.11 Jumlah UMKM Makanan Minuman yang Mendaftar di Tokopedia Berdasarkan Tahun.	104
Gambar 6.12 Pratinjau Data Kategori Makanan dan Minuman Shopee.	104
Gambar 6.13 Pratinjau Data Produk Makanan dan Minuman Shopee.	105
Gambar 6.14 Jumlah Data Tiap Kategori Makanan dan Minuman Shopee.....	106
Gambar 6.15 <i>Wordcloud</i> Produk Makanan Minuman Shopee	106
Gambar 6.16 Pratinjau Data <i>URL</i> UMKM Makanan dan Minuman Shopee.....	107
Gambar 6.17 Pratinjau Data UMKM Shopee Hasil <i>Scraping</i>	108
Gambar 6.18 Pratinjau Data UMKM Shopee Hasil <i>Preprocess</i>	109
Gambar 6.19 Perbandingan Data UMKM Shopee Berdasarkan Tipe <i>Seller</i>	110
Gambar 6.20 Perbandingan Data UMKM Shopee Berdasarkan Rata-Rata <i>Rating</i>	111
Gambar 6.21 Perbandingan Data UMKM Shopee Berdasarkan Rata-Rata <i>Review</i>	112
Gambar 6.22 Pratinjau Data UMKM Google My Business Hasil <i>Scraping</i>	112
Gambar 6.23 Pratinjau Data UMKM Google My Business Hasil <i>Preprocess</i>	113
Gambar 6.24 Laman Visualisasi Informasi.	115
Gambar 6.25 Halaman <i>General Comparison</i>	116
Gambar 6.26 Halaman Visualisasi Data UMKM Lokal.....	117
Gambar 6.27 Halaman Visualisasi Data UMKM Tokopedia.	117
Gambar 6.28 Halaman Visualisasi Data UMKM Shopee. ...	118

DAFTAR TABEL

Tabel 2.1 Pekerjaan Terkait.	7
Tabel 2.2 Kriteria UMKM.	9
Tabel 3.1 Rangkuman Metodologi.	23
Tabel 4.1 Rancangan Interaksi Elemen Halaman Awal Tokopedia.	26
Tabel 4.2 Rancangan Data Kategori Tokopedia.	27
Tabel 4.3 Rancangan Lokasi Elemen Data Kategori Tokopedia.	27
Tabel 4.4 Rancangan Interaksi Elemen Halaman Kategori Produk Tokopedia.	28
Tabel 4.5 Rancangan Data Produk Tokopedia.	28
Tabel 4.6 Rancangan Lokasi Elemen Data Produk Tokopedia.	29
Tabel 4.7 Jenis <i>URL</i> Produk di Tokopedia.	29
Tabel 4.8 Rancangan Data dan <i>Preprocess URL</i> UMKM Tokopedia.	30
Tabel 4.9 Rancangan Interaksi Elemen Halaman UMKM Tokopedia.	31
Tabel 4.10 Rancangan Data UMKM Tokopedia.	32
Tabel 4.11 Rancangan Lokasi Elemen Data UMKM Tokopedia.	33
Tabel 4.12 Rancangan Interaksi Elemen Halaman Kategori Produk Shopee.	36
Tabel 4.13 Rancangan Data Kategori Produk Shopee.	36
Tabel 4.14 Rancangan Lokasi Elemen Data Kategori Produk Shopee.	37
Tabel 4.15 Rancangan Interaksi Elemen Halaman Kategori Produk Shopee.	38
Tabel 4.16 Rancangan Data Produk Shopee.	39
Tabel 4.17 Rancangan Lokasi Elemen Data Produk Shopee.	39
Tabel 4.18 Rancangan Data <i>URL</i> UMKM Shopee.	40
Tabel 4.19 Rancangan Lokasi Elemen Data Produk Shopee.	40
Tabel 4.20 Rancangan Data UMKM Shopee.	40
Tabel 4.21 Rancangan Lokasi Elemen Data UMKM Shopee.	41
Tabel 4.22 Rancangan Data Kategori Google My Business.	44
Tabel 4.23 Rancangan Data Kecamatan dan Kelurahan.	44

Tabel 4.24 Rancangan Interaksi Google My Business.....	45
Tabel 4.25 Rancangan Data UMKM Google My Business. ...	46
Tabel 4.26 Rancangan Lokasi Elemen Data UMKM Google My Business.....	47
Tabel 4.27 Rancangan Interaksi Elemen Pada Instagram.	49
Tabel 4.28 Rancangan Data Hasil Pencarian Pada Instagram.	49
Tabel 4.29 Rancangan Lokasi Elemen Data Hasil Pencarian Pada Instagram.	50
Tabel 4.30 Rancangan <i>Data Integration</i>	51
Tabel 4.31 Rancangan Matriks Visualisasi.	54
Tabel 4.32 Rancangan Visualisasi Data.	56
Tabel 5.1 Implementasi XPath Elemen Lokasi Data Kategori Tokopedia.....	61
Tabel 5.2 Implementasi XPath Lokasi Elemen Data Produk Tokopedia.....	63
Tabel 5.3 Implementasi XPath Lokasi Elemen Data UMKM Tokopedia.....	66
Tabel 5.4 Implementasi XPath Lokasi Elemen Data Kategori Produk Shopee.....	69
Tabel 5.5 Implementasi XPath Lokasi Elemen Data Produk Shopee.	71
Tabel 5.6 Rancangan Lokasi Elemen Data Produk Shopee. ..	71
Tabel 5.7 Implementasi XPath Lokasi Elemen Data UMKM Shopee.	72
Tabel 5.8 Implementasi XPath Elemen Lokasi Data UMKM Google My Business.	78
Tabel 5.9 <i>Preprocess</i> Data UMKM Tokopedia.....	81
Tabel 5.10 <i>Preprocessing</i> Data Shopee.....	83
Tabel 5.11 <i>Preprocess</i> Data UMKM Google My Business. ...	85
Tabel 5.12 <i>Threshold</i> Nilai <i>Similarity</i> Tiap Atribut.	88
Tabel 5.13 Hasil Pengujian.	89
Tabel A.1 Kode Program <i>Scraping</i> Kategori Tokopedia.	129
Tabel A.2 Kode Program <i>Scraping</i> Produk Tokopedia.....	129
Tabel A.3 Kode Program <i>Preprocess URL</i> Produk Menjadi URL UMKM Tokopedia.	129
Tabel A.4 Kode Program <i>Scraping</i> Informasi UMKM Makanan dan Minuman Tokopedia.....	129

Tabel A.5 Kode Program <i>Scraping</i> Kategori Makanan dan Minuman Shopee.	130
Tabel A.6 Kode Program <i>Scraping</i> Produk Makanan dan Minuman Shopee.	130
Tabel A.7 Kode Program <i>Scraping URL</i> UMKM Makanan dan Minuman Shopee.	130
Tabel A.8 Kode Program <i>Scraping</i> Informasi UMKM Makanan dan Minuman Shopee.....	130
Tabel A.9 Kode Program <i>Scraping</i> Informasi UMKM Makanan dan Minuman Google My Business.	131
Tabel A.10 Kode Program <i>Preprocess</i> Data UMKM Makanan dan Minuman Tokopedia.	131
Tabel A.11 Kode Program <i>Preprocess</i> Data UMKM Makanan dan Minuman Shopee.....	132
Tabel A.12 Kode Program <i>Preprocess</i> Data UMKM Makanan dan Minuman Google My Business.	132
Tabel A.13 Kode Program Integrasi Data UMKM Makanan dan Minuman Tokopedia dengan Shopee.....	132
Tabel A.14 Kode Program Integrasi Data UMKM Makanan dan Minuman Google My Business dengan Tokopedia dan Shopee.	133
Tabel B.1 Data Kategori Makanan dan Minuman Tokopedia.	134
Tabel B.2 Data Produk Makanan dan Minuman Tokopedia.....	134
Tabel B.3 Data <i>URL</i> UMKM Makanan dan Minuman Tokopedia.....	134
Tabel B.4 Data UMKM Makanan dan Minuman Tokopedia.	134
Tabel B.5 Data Kategori Makanan dan Minuman Shopee... ..	135
Tabel B.6 Data Produk Makanan dan Minuman Shopee.	135
Tabel B.7 Data <i>URL</i> UMKM Makanan dan Minuman Shopee.	135
Tabel B.8 Data UMKM Makanan dan Minuman Shopee....	135
Tabel B.9 Data UMKM Makanan dan Minuman Tokopedia Hasil <i>Preprocess</i>	135
Tabel B.10 Data UMKM Makanan dan Minuman Shopee Hasil <i>Preprocess</i>	136

Tabel B.11 Data UMKM Makanan dan Minuman Google My Business Hasil <i>Preprocess</i>	136
Tabel B.12 Data UMKM Makanan dan Minuman Hasil Integrasi Tokopedia dengan Shopee.....	136
Tabel B.13 Data UMKM Makanan dan Minuman Hasil Integrasi Google My Business dengan Tokopedia dan Shopee.	137
Tabel C.1 Tabel <i>Chi Square</i>	139

BAB I

PENDAHULUAN

Bab ini akan menjelaskan tentang pendahuluan pengerjaan tugas akhir yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat yang akan diperoleh dari penelitian tugas akhir ini.

1.1 Latar Belakang

Usaha Mikro, Kecil dan Menengah (UMKM) mempunyai peran penting dan strategis dalam pembangunan ekonomi nasional. Selain berperan dalam pertumbuhan ekonomi dan penyerapan tenaga kerja, UMKM juga berperan dalam mendistribusikan hasil-hasil pembangunan. Usaha Mikro, Kecil dan Menengah (UMKM) memiliki peranan penting dalam perekonomian di Indonesia. UMKM memiliki proporsi sebesar 99,99% dari total keseluruhan pelaku usaha di Indonesia atau sebanyak 56,54 juta unit [1]. Bisnis UMKM menyumbang PDB (Produk Domestik Bruto) sekitar 60% dan membuka lapangan pekerjaan bagi masyarakat [1].

Dengan meningkatnya aktifitas ekonomi dan perkembangan internet, peluang pengembangan UMKM meningkat. Salah satu peluang yang dapat diambil adalah dengan memanfaatkan platform daring. Platform daring memberikan peluang bagi UMKM untuk meningkatkan peluang pasar. Peluang yang dapat dimiliki dengan mengadopsi platform daring adalah peluang akses terhadap UMKM melalui *multi-platform* atau banyak platform. Menurut Indonesia Millennial Report 2019 [2], salah satu pengguna internet terbanyak, yaitu milenial sangat suka mencari perbandingan harga, fitur, program promo dan kualitas produk di beberapa platform pasar *online* sebelum memutuskan membeli sebuah barang. Hal tersebut menunjukkan akses *multi-platform* menjadi salah satu opsi untuk menjangkau UMKM.

Namun, hasil survei pasar *online*, salah satu bentuk platform daring, yang dilakukan oleh Badan Pusat Statistik (BPS) pada tahun 2019 dengan sampel sebanyak 3504 usaha yang dilakukan pendataan mengatakan bahwa hanya 15,08 persen yang

merupakan usaha dalam pasar *online* atau sebanyak 528 dari 3504 usaha [3].

Salah satu industri yang memiliki potensi untuk memanfaatkan data adalah industri makanan dan minuman. Dilihat dari jumlah pada tahun 2010, jumlah usaha makanan dan minuman di kategori industri kecil dan mikro mencapai 960.305 perusahaan dan meningkat hingga mencapai 1.614.149 perusahaan di tahun 2015. Rata-rata pertumbuhan jumlah usaha dari industri makanan dan minuman di kategori usaha mikro dan kecil pada periode 2010 sampai 2015 mencapai 11,61 persen [4].

Industri makanan dan minuman menjadi salah satu sektor unggulan dari sektor manufaktur Indonesia. Tingginya nilai dari industri makanan dan minuman menyebabkan sektor ini berkontribusi besar terhadap PDB dari Indonesia. Nilai dari Industri makanan dan minuman menunjukkan tren yang meningkat dari tahun 2014 sampai tahun 2018. Pada tahun 2014 nilai dari industri makanan telah mencapai 117 triliun rupiah, dan di tahun 2018 sektor industri makanan dan minuman telah meningkat dan mencapai 165 triliun rupiah [4].

Pemanfaatan *e-commerce* sebagai salah bentuk platform daring pada industri makanan minuman juga terjadi di Amerika Serikat. Data sensus Amerika Serikat menunjukkan bahwa pada tahun 2012 nilai *e-commerce* pengiriman produk makanan menyumbang lebih dari 50 persen dari total nilai pendapatan (diperkirakan sekitar 750 miliar USD) [5]. Salah satu pengembangan platform *e-commerce*, yaitu eBay, memberikan peluang bagi bisnis makanan kecil atau bahkan individu untuk menjual produk makanan secara online [5].

Dalam menemukan peluang *e-commerce* perlu untuk mengetahui kondisi saat ini. Salah satu cara untuk mengetahui kondisi saat ini adalah dengan melakukan analisis data. Analisis data eksisting adalah cara yang hemat biaya untuk memanfaatkan data yang sudah dikumpulkan untuk menjawab pertanyaan-pertanyaan penelitian atau untuk memberikan penilaian yang lebih dari studi sebelumnya [6].

Hasil dari tugas akhir ini berupa data UMKM makanan dan minuman, dan dokumen hasil analisis kondisi pemanfaatan platform daring pada UMKM makanan dan minuman di Surabaya. Harapannya hasil dari tugas akhir ini dapat digunakan oleh UMKM dan masyarakat untuk mengetahui kondisi UMKM makanan dan minuman pada kanal daring, dan sebagai *channel* dalam menjangkau UMKM makanan dan minuman di Surabaya bagi pelanggan.

1.2 Rumusan Masalah

UMKM menjadi salah satu bisnis yang membutuhkan akses terhadap data, dan informasi untuk mengetahui kondisi bisnis saat ini. Data tidak hanya dibutuhkan oleh pelaku bisnis, melainkan dibutuhkan juga oleh pelanggan. Dengan adanya data, pelanggan memiliki kemudahan dalam mengetahui cara menjangkau UMKM dalam platform daring yang berbeda-beda. Ketersediaan data yang dapat diakses oleh pelaku usaha maupun pelanggan terbatas pada platform-platform daring, sehingga data tidak tersedia secara menyeluruh disebabkan karena keterbatasan akses. Keterbatasan akses pada data menyebabkan tidak dapat dilakukannya analisis mengenai kondisi UMKM makanan dan minuman di Surabaya dalam penerapan teknologi informasi platform daring. Masalah yang terjadi dapat diringkas menjadi tiga poin, yaitu

1. Bagaimana menyediakan data UMKM makanan dan minuman antar platform daring?
2. Bagaimana kondisi data UMKM makanan dan minuman di Surabaya pada platform daring?
3. Bagaimana kondisi pemanfaatan platform daring oleh UMKM makanan dan minuman di Surabaya?

1.3 Batasan Permasalahan

Batasan masalah pada tugas akhir ini sebagai berikut.

1. Data yang dikumpulkan merupakan data UMKM kategori makanan dan minuman di Surabaya.

2. Data didapatkan dari platform Google My Business, Tokopedia, Shopee, dan Instagram.
3. Data yang didapat digunakan untuk keperluan *descriptive analytics* atau analisis deskriptif.

1.4 Tujuan

Tujuan dari pembuatan tugas akhir ini sebagai berikut:

1. Mengidentifikasi data UMKM melalui observasi *website* yang ada untuk menghasilkan skema data UMKM makanan dan minuman di Surabaya dari berbagai platform daring.
2. Melakukan pengumpulan dan pengolahan data melalui protokol *web* untuk menghasilkan kumpulan data UMKM makanan dan minuman di Surabaya dari berbagai platform daring.
3. Mengidentifikasi kondisi pemanfaatan platform daring pada UMKM makanan dan minuman di Surabaya dalam bentuk visualisasi informasi, dan analisis terhadap hasil visualisasi.

1.5 Metode

Metode yang digunakan pada penelitian ini, yaitu *web scraping*; *Extraction, Transformation, and Loading (ETL)*; *entity resolution*, dan analisis deskriptif.

1.6 Manfaat

Berikut manfaat yang diharapkan dari pengerjaan tugas akhir ini:

1. Secara Keilmuan

Skema dan kumpulan data yang dihasilkan dapat digunakan sebagai bahan, dan referensi untuk melakukan analisis lebih lanjut mengenai UMKM kategori makanan dan minuman di Surabaya. Kode pengumpulan data yang berhasil dibuat dapat digunakan sebagai *template*, referensi, atau acuan dalam melakukan

pengumpulan data pada platform daring yang digunakan.

2. Secara Praktis

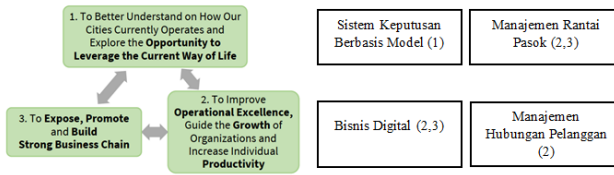
Visualisasi informasi sebagai penggambaran kondisi pemanfaatan platform daring oleh UMKM makanan dan minuman di Surabaya, sehingga dapat menjadi informasi pendukung dalam pengambilan keputusan yang berkaitan dengan kondisi UMKM kategori makanan dan minuman di Surabaya dalam memanfaatkan platform daring.

1.7 Target Luaran

Target luaran dalam penelitian ini dikategorikan menjadi aspek teknis dan dokumen. Dalam aspek teknis ditargetkan luaran berupa kode pengumpulan data, skema data, kumpulan data UMKM makanan dan minuman pada beberapa platform daring, dan visualisasi informasi serta analisis. Dalam aspek dokumen ditargetkan luaran berupa dokumen laporan tugas akhir, dan jurnal mahasiswa ITS.

1.8 Relevansi

Tugas akhir ini termasuk dalam topik pengetahuan ‘*Bisnis Digital*’ yang berkontribusi pada tujuan penelitian Laboratorium Sistem Enterprise yaitu untuk mengetahui dan lebih memahami keadaan bisnis saat ini sehingga dapat memandu pertumbuhan organisasi dan produktivitas individu; Untuk mengekspos, mempromosikan, dan membangun rantai bisnis yang kuat. Pada Gambar 1.1 digambarkan tujuan penelitian dan topik pengetahuan di Laboratorium Sistem Enterprise.



Gambar 1.1 Kerangka Kerja Riset Laboratorium Sistem *Enterprise*.

BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka terdiri dari landasan-landasan yang akan digunakan dalam penelitian tugas akhir ini. Adapun cakupan dari tinjauan pustaka meliputi penelitian-penelitian sebelumnya, kajian pustaka, dan metode yang digunakan selama pengerjaan.

2.1 Penelitian Sebelumnya

Terdapat beberapa penelitian yang memiliki topik yang hampir serupa dengan penelitian ini. Adapun penelitian-penelitian tersebut dijelaskan pada Tabel 2.1.

Tabel 2.1 Pekerjaan Terkait.

Penelitian 1	
Judul	<i>Trends and Applications in Knowledge Discovery and Data Mining</i> [7]
Gambaran umum	Sebuah buku yang menunjukkan metode untuk melakukan <i>entity resolution</i> atau menemukan sebuah entitas yang sama dari sekumpulan data. Metode yang digunakan adalah dengan melihat kemiripan data (nama) produk berjenis teks antara satu platform <i>e-commerce</i> dengan <i>e-commerce</i> lain.
Keterkaitan penelitian	<i>Tools</i> yang digunakan sama yaitu <i>sequence matcher</i> atau merupakan salah satu <i>library</i> yang terdapat pada bahasa pemrograman <i>python</i> . <i>Tools</i> akan digunakan untuk menentukan entitas (UMKM) yang sama antar platform kanal daring.
Penelitian 2	
Judul	<i>Web Scraping Data Extraction from Websites</i> [8]
Gambaran umum	Penelitian yang memberikan pemahaman menyeluruh terkait dengan <i>web scraping</i> ,

	mulai dari definisi, jenis-jenis, metode, hingga proses ekstraksi informasi dari data yang dikumpulkan.
Keterkaitan penelitian	Metode dan <i>tools</i> yang digunakan untuk melakukan proses <i>web scraping</i> . <i>Scraping</i> yang dilakukan menggunakan metode DOM <i>parsing</i> mulai dari mendefinisikan kebutuhan, persiapan, hingga eksekusi <i>script</i> .
Penelitian 3	
Judul	<i>Comparison of E-commerce Products using web mining</i> [9]
Gambaran umum	Ekstraksi informasi dan melakukan analisis komparasi produk dari hasil <i>web scraping</i> .
Keterkaitan penelitian	Analisis komparasi dari entitas UMKM dengan memperhatikan metrik-metrik tertentu.
Penelitian 4	
Judul	<i>From Price Collection to Price Data Analytics How New Large Data Sources Require Price Statisticians to Re-Think Their Index Compilation Procedures Experiences From Web-Scraped and Scanner Data</i> [10]
Gambaran umum	Penelitian dengan proses mengubah data kumpulan harga menjadi analisis data harga. Terdapat proses integrasi multi-platform.
Keterkaitan penelitian	Mengubah data hasil <i>scraping</i> menjadi sebuah analisis tentang kondisi suatu entitas.
Penelitian 5	

Judul	<i>Cloud Based Web Scraping for Big Data Applications</i> [11]
Gambaran umum	Penelitian ini mengembangkan metode dalam melakukan <i>web scraping</i> . <i>Web scraping</i> yang dilakukan pada penelitian ini dikhususkan pada <i>big data</i> . Pada penelitian ini juga diberikan penjelasan hasil dari implementasi <i>web scraping</i> . Hasil pengukuran dibandingkan terdiri dari beberapa <i>tools scraper</i> yang digunakan.
Keterkaitan Penelitian	Melakukan <i>web scraping</i> pada sebuah <i>website</i> . Pada penelitian ini spesifik implementasi <i>cloud based</i> untuk <i>big data</i> . Secara garis besar metode yang akan digunakan untuk pembuatan <i>scraper</i> akan sama, tetapi diperlukan kesuaian untuk penggunaan <i>website e-commerce</i> .

2.2 Usaha Mikro Kecil Menengah (UMKM)

Di Indonesia, terdapat berbagai definisi yang berbeda mengenai UMKM berdasarkan kepentingan lembaga yang memberi definisi. Menurut Undang-Undang No. 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah [12] menetapkan beberapa kriteria UMKM seperti yang tertera pada Tabel 2.2.

Tabel 2.2 Kriteria UMKM.

UMKM	Kriteria Kekayaan Bersih (rupiah)	Kriteria Penjualan (rupiah)
Usaha Mikro	0 – 50 juta	0 – 300 juta
Usaha Kecil	50 – 500 juta	300 juta – 2,5 milyar

Usaha Menengah	500 juta – 10 milyar	2,5 – 50 milyar
----------------	----------------------	-----------------

Usaha Mikro memiliki kekayaan bersih paling banyak Rp. 50.000.000 (lima puluh juta rupiah) (tidak termasuk tanah dan bangunan), atau memiliki hasil penjualan tahunan paling banyak Rp. 300.000.000 (tiga ratus juta rupiah). Usaha Kecil memiliki kekayaan bersih lebih dari Rp. 50.000.000 (lima puluh juta rupiah) hingga sampai paling banyak Rp. 500.000.000 (lima ratus juta rupiah) tidak termasuk tanah dan bangunan tempat usaha, atau memiliki hasil penjualan tahunan lebih dari Rp. 300.000.000 (tiga ratus juta rupiah) sampai dengan paling banyak Rp. 2.500.000.000 (dua miliar lima ratus juta rupiah). Usaha Menengah memiliki kekayaan bersih lebih dari Rp. 500.000.000 (lima ratus juta rupiah) sampai dengan paling banyak Rp. 10.000.000.000 (sepuluh miliar rupiah) tidak termasuk tanah dan bangunan tempat usaha, atau memiliki hasil penjualan tahunan lebih dari Rp. 2.500.000.000,00 (dua miliar lima ratus juta rupiah) sampai dengan paling banyak Rp. 50.000.000.000 (lima puluh miliar rupiah).

2.3 Data Collection

Dalam proses pengolahan data, *data collection* atau proses mengumpulkan data menjadi hal yang penting. *Data collection* merupakan sebuah proses untuk mengumpulkan data dan mendapatkan informasi. Tujuan untuk pengumpulan data adalah untuk menangkap bukti yang berbentuk informasi kemudian diterjemahkan menjadi analisis data. *Data collection* memungkinkan untuk mendukung jawaban atas pertanyaan yang diajukan. Metode dalam melakukan *data collection* bervariasi, tetapi tetap menekankan pada pengumpulan data yang akurat dan jujur. *Data collection* yang akurat sangat penting untuk menjaga integritas penelitian [13].

2.4 Web Scraping

Web scraping merupakan sebuah praktik mengumpulkan data melalui *web browser*. Secara umum, *web scraping* dilakukan

dengan menuliskan program otomatis yang mengirimkan permintaan kepada *website*, meminta data (biasanya dalam bentuk HTML), kemudian mengekstraksi data yang dibutuhkan [14]. Dengan kata lain, kita dapat mengatakan bahwa daripada secara manual menyimpan data dari situs web, *web scraping* akan secara otomatis memuat dan mengekstrak data dari *website* sesuai kebutuhan. Metode yang digunakan untuk *web scrapig* adalah *Document Object Model (DOM) Parsing*, merupakan evolusi dari *HTML Parsing* berdasarkan perkembangan bahasa dan *browser* yang mengarah pada pengenalan *Document Object Model*. DOM banyak digunakan untuk *Cascading Stylesheets (CSS)* dan *JavaScript* [8].

2.5 *Extraction, Transformation, Loading (ETL)*

Extraction, Transformation, Loading (ETL) merupakan proses dalam mengintegrasikan data. Dalam prosesnya melibatkan ekstraksi (*extraction*) data dari sumber luar sistem, merubahnya (*transformation*) menjadi sesuai kebutuhan bisnis, dan kemudian memuatnya (*loading*) ke dalam *data warehouse* atau sistem yang digunakan [15].

2.6 *Entity Resolution*

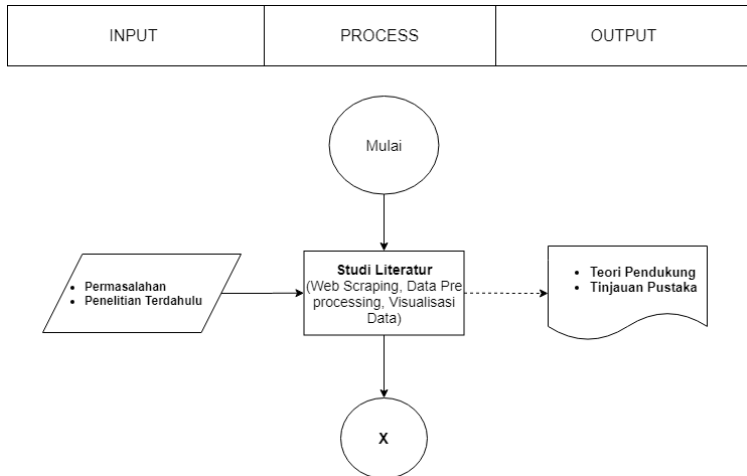
Entity resolution mendeskripsikan identifikasi, ekstraksi, dan mencocokkan entitas yang sama dalam data yang terstruktur maupun tidak terstruktur dari satu atau berbagai sumber yang berbeda [7]. Istilah *entity resolution* digunakan untuk melakukan integrasi data dengan mencocokkan entitas yang sama antar platform yang berbeda.

BAB III METODOLOGI

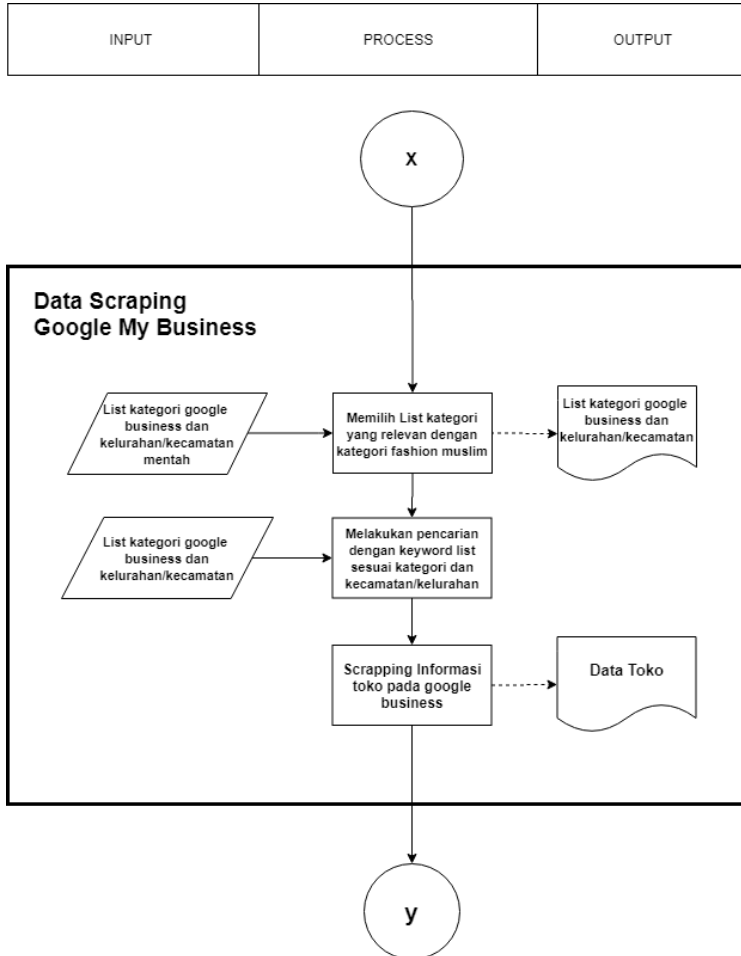
Pada bab ini dijelaskan metodologi yang digunakan sebagai panduan untuk menyelesaikan penelitian tugas akhir ini. Metodologi berdampak pada proses pengerjaan yang terstruktur dan terarah.

3.1 Diagram Metodologi

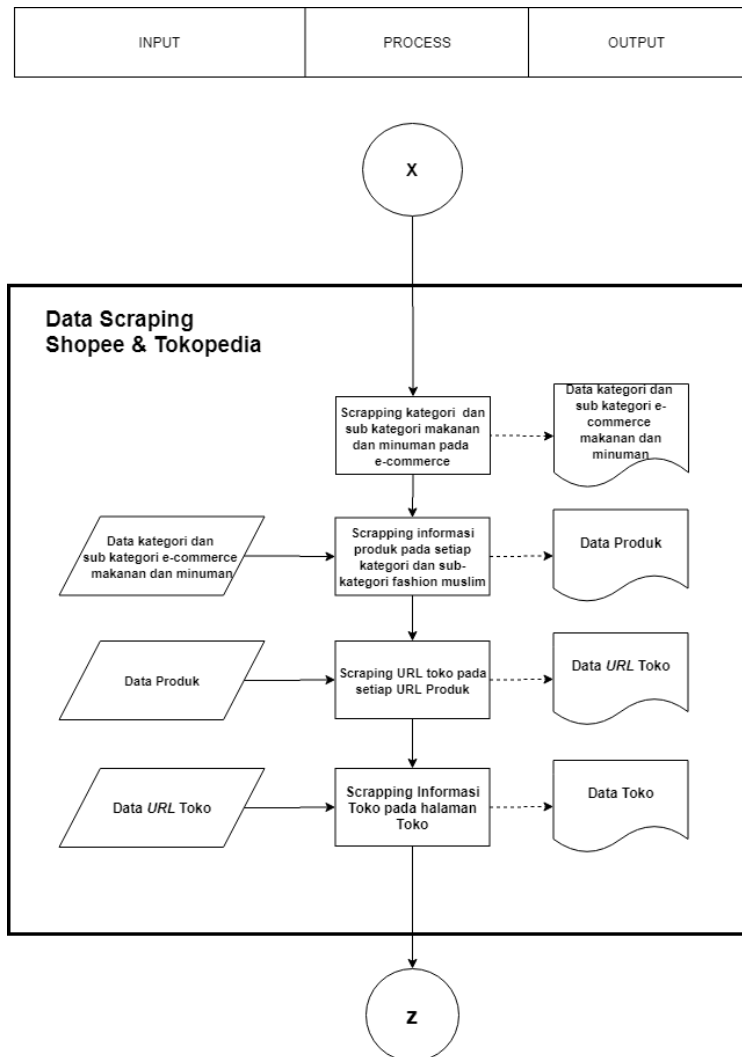
Sub-bab ini menjelaskan tahapan dan proses yang dilakukan dalam penelitian ini. Rangkaian tahapan dan prosesnya terlihat dalam metodologi penelitian yang disusun secara berurutan pada Gambar 3.1, Gambar 3.2, Gambar 3.3, Gambar 3.4, dan Gambar 3.5.



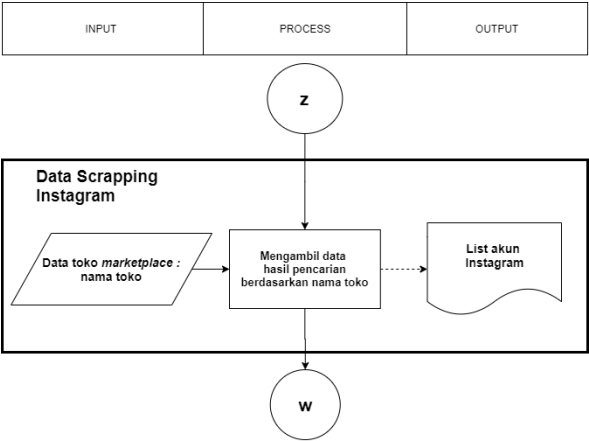
Gambar 3.1 Metodologi Penelitian Bagian 1.



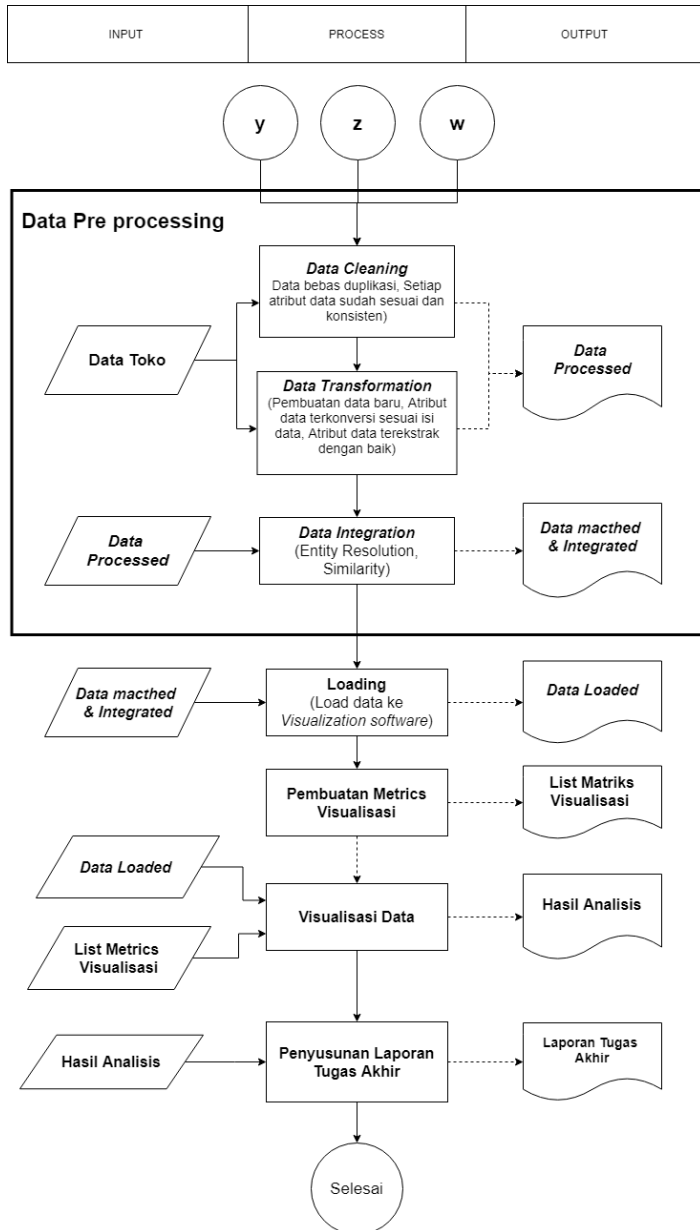
Gambar 3.2 Metodologi Penelitian Bagian 2.



Gambar 3.3 Metodologi Penelitian Bagian 3.



Gambar 3.4 Metodologi Penelitian Bagian 4.



Gambar 3.5 Metodologi Penelitian Bagian 5.

3.2 Uraian Metodologi

Bagian ini merupakan penjelasan-penjelasan uraian dari metodologi pengerjaan tugas akhir. Adapun uraian tersebut sebagai berikut.

3.2.1 Studi Literatur

Studi literatur merupakan tahapan paling awal yang dilakukan pada penulisan tugas akhir ini yang bertujuan untuk menentukan topik dengan menganalisis latar belakang permasalahan terkait perlunya analisis kondisi pemanfaatan kanal daring pada UMKM makanan dan minuman di Surabaya. Literatur yang dikaji terkait dengan topik yang diangkat, yaitu kondisi UMKM secara umum, metode pengumpulan data menggunakan *web scraping*, metode *pre-process*, integrasi data, dan visualisasi terkait data UMKM. Pada proses ini ditentukan batasan masalah, tujuan, serta manfaat dari penelitian.

3.2.2 Data Scraping

Setelah mengkaji literatur dan penelitian terkait, proses berikutnya adalah pengumpulan data. Data yang digunakan pada penelitian ini adalah data yang didapat dari *scraping* pada tiga platform kanal daring, yaitu Google My Business, Tokopedia, dan Shopee.

1. *Scraping* data Tokopedia dan Shopee

Data scraping untuk mendapatkan informasi UMKM pada platform *marketplace* yaitu Tokopedia dan Shopee membutuhkan beberapa data seperti *URL* kategori, informasi produk, *URL* produk dan *URL* UMKM. Adapun proses *data scraping* pada platform *e-commerce* terdiri dari beberapa tahap, yaitu :

a. *Scraping* kategori dan sub kategori makanan dan minuman

Kategori yang dipilih adalah kategori makanan dan minuman. Sebelum melakukan pengambilan data, perlu dilakukan pemahaman struktur halaman *website* dan lokasi elemen data yang akan diambil. Hasil dari proses ini adalah data kategori beserta sub kategori.

b. *Scraping* informasi produk

Tahap kedua *data scraping* pada Tokopedia dan Shopee adalah mengambil informasi produk. Hasil dari proses ini adalah list data produk dan URL produk.

2. *Scraping* Google My Business

Kegiatan *data scrapping* untuk mendapatkan informasi UMKM pada platform Google My Business. Adapun proses *data scrapping* pada platform Google My Business terdiri dari beberapa tahap, yaitu :

a. Mengumpulkan data pendukung

Tahap pertama *data scrapping* pada platform Google My Business adalah mengumpulkan data pendukung. Data tersebut merupakan data kategori Google My Business dan data lokasi yang didapatkan dari sumber di internet. Adapun hasil dari aktivitas ini adalah data kategori Google My Business dan data lokasi.

b. Memilih kategori yang relevan dengan makanan dan minuman

Tahap kedua *scraping* data pada platform google business adalah memilih list kategori yang relevan dengan kategori makanan dan minuman. Pemilihan kategori dilakukan secara manual dan *semi-preprocessing* . Kategori yang dianggap tidak relevan akan dibuang dan tersisa kategori yang relevan dengan kategori makanan dan minuman. Hasil dari aktivitas ini adalah list data kategori google business relevan dengan kategori makanan dan minuman.

c. *Scraping* informasi UMKM pada Google My Business

Tahap ketiga *scraping data* pada platform Google My Business adalah *scraping* informasi UMKM. Pada tahap ini diperlukan data masukan dari data pendukung. Penggunaan data tersebut digunakan

untuk membuat kata kunci pencarian yang akan digunakan pada pencarian UMKM di Google My Business. Hasil pencarian setiap UMKM yang muncul akan diambil informasi nya Hasil dari aktivitas ini adalah data UMKM Google My Business.

3. *Scrapping data Instagram*

Kegiatan *scraping* di Instagram menggunakan data UMKM yang didapat dari Tokopedia dan Shopee yaitu nama UMKM. Kegiatan ini dengan tujuan mengambil data akun UMKM berdasarkan nama UMKM. Hasil dari aktivitas ini adalah data akun instagram.

3.2.3 *Pre-processing*

Pada tahap *pre-processing*, digunakan juga bahasa python. Penggunaan python dimaksudkan untuk mengotomatis proses secara menyeluruh terhadap data.

A. *Extraction*

Pada tahap *extraction* data yang berhasil dikumpulkan melalui *web scraping* diekstrak dengan menambahkan beberapa atribut tambahan. Atribut tambahan digunakan untuk meningkatkan kekayaan data dengan menambahkan variasi atribut dari yang telah didapatkan.

B. *Transformation*

Pada tahap *transformation* terdapat beberapa hal yang dilakukan, yaitu menghapus duplikasi data, standardisasi atribut dengan mengubah jenis atau pun *value* dari tiap atribut agar memiliki keseragaman, menghapus data yang tidak relevan dengan kategori makanan dan minuman, dan menghapus data UMKM yang berada di luar surabaya.

C. *Integration*

Pada tahap *integration* menggunakan konsep *entity resolution*. *Entity resolution* dilakukan dengan menggunakan teknik *similarity* atau melihat kemiripan antar entitas UMKM dalam data. Dalam melakukan *entity*

resolution diperlukan adanya *threshold* atau ambang batas minimum hasil *similarity* sehingga dianggap baik. *Threshold* didapatkan dengan cara menghitung banyaknya jumlah benar atau *True Positive* dari hasil *similarity*. Teknik *similarity* dilakukan dengan menggunakan *library sequencematcher* pada python. Dari hasil *similarity* yang muncul, diambil hasil yang paling besar untuk tiap UMKM. Dari tiap hasil maksimal diambil lagi berdasarkan *threshold* atau ambang batas minimum yang telah ditentukan. Hasil *similarity* maksimal yang memiliki nilai lebih dari atau sama dengan *threshold* maka dianggap sebagai *match* atau cocok.

3.2.4 Data Loading

Pada tahap *loading*, data yang telah diintegrasikan dimasukkan ke dalam *software* analisis. *Software* analisis yang digunakan adalah Tableau. Dari hasil integrasi, data-data antar platform akan digabungkan menjadi satu sesuai dengan hasil integrasi yang didapatkan. Penggabungan data dilakukan dengan tujuan agar analisis dapat dilakukan secara lintas platform.

3.2.5 Pembuatan Matriks Visualisasi

Sebelum data diubah menjadi informasi visual, data dapat dilakukan analisis *metrics* untuk mengetahui informasi apa yang dapat ditampilkan dari data UMKM dan apa manfaat dari hasil informasi visual tersebut. Hasil dari aktivitas ini adalah tujuan dibuatnya tiap visualisasi.

3.2.6 Visualisasi dan Analisis

Pada tahap visualisasi dan analisis, data hasil penggabungan diolah menjadi tampilan-tampilan visualisasi informasi, seperti diagram. Analisis yang dilakukan terhadap hasil visualisasi informasi berupa analisis deskriptif sesuai dengan metrik yang telah ditentukan. Analisis dilakukan secara lintas platform pada data yang memiliki kecocokan lintas platform. Hasil dari proses analisis adalah kesimpulan terhadap kondisi data atau pun UMKM makanan dan minuman di Surabaya.

3.2.7 Pelaporan

Pada tahap pelaporan, semua hasil perancangan, implementasi, hasil analisis, serta kesimpulan dimasukkan ke dalam dokumentasi. Laporan yang dihasilkan dalam penelitian ini adalah buku tugas akhir dan jurnal mahasiswa ITS.

3.3 Rangkuman Metodologi

Rangkuman metodologi berisikan mengenai metodologi yang dilakukan dalam penelitian ini, dimulai dari rangkaian aktifitas, tujuan, input, output dan metode yang digunakan seperti pada Tabel 3.1.

Tabel 3.1 Rangkuman Metodologi.

Aktivitas	Tujuan	Input	Output	Metode
Studi Literatur	Memahami domain masalah dan teknis penyelesaian masalah	Penelitian sebelumnya	Pemahaman masalah dan penyelesaian masalah,	Observasi penelitian terkait
Pengembangan <i>scraper</i>	Penyediaan alat untuk pengambilan data	Daftar kebutuhan data dan pemahaman struktur halaman <i>website</i>	<i>Data scraper</i>	<i>Web scraper</i> menggunakan referensi pekerjaan terkait sebelumnya
Pengambilan data	Mengumpulkan data yang dibutuhkan untuk dianalisis	Data pendukung, dan <i>data scraper</i>	Data UMKM Tokopedia, Shopee, dan Google My Business	Menjalankan <i>data scraper</i> yang telah dibuat sesuai dengan platform (Tokopedia, Shopee, dan Google My Business)
Pengolahan data	Mengolah data agar siap untuk dianalisis	Data UMKM Tokopedia, Shopee, dan Google My Business	Data siap analisis	<i>Data cleaning, extraction, transformation</i>
Integrasi Data	Menggabungkan data antar platform	Data UMKM Tokopedia, Shopee,	Data terintegrasi	<i>Entity resolution</i>

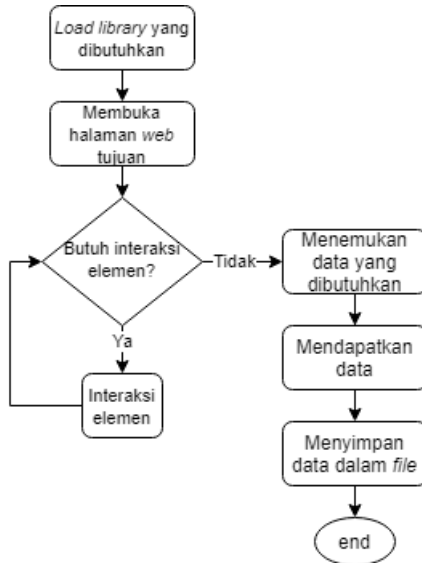
Aktivitas	Tujuan	Input	Output	Metode
		dan Google My Business		
Pembuatan matriks visualisasi	Menemukan matriks atau <i>measure</i> yang dapat memberikan <i>insight</i> terhadap data	Data terintegrasi, data UMKM Tokopedia, Shopee, dan Google My Business	Daftar matriks visualisasi	Menentukan matriks berdasarkan hasil analisis data yang berhasil dikumpulkan
Visualisasi Informasi	Menampilkan visualisasi terhadap data sehingga dapat memberikan informasi	Data terintegrasi, data UMKM Tokopedia, Shopee, dan Google My Business	Visualisasi data	Menggunakan <i>tools</i> visualisasi data
Penyusunan Laporan	Menghasilkan laporan hasil pengumpulan, pengolahan, dan analisis data	Data UMKM semua platform, <i>scraper</i> , visualisasi, dan analisis data	Laporan hasil analisis data	Penulisan laporan

BAB IV PERANCANGAN

Pada bab ini akan dibahas mengenai seluruh proses perancangan *scraping* data, *pre-process* data, serta visualisasi dan analisis data.

4.1 Rancangan Data Scraper

Scraping *data* dibagi menjadi beberapa platform, yaitu Tokopedia, Shopee, dan Google My Business. Data-data yang diperlukan adalah data-data terkait dengan proses, seperti data *URL*, lokasi elemen data, dan daftar interaksi dengan elemen yang diperlukan untuk memunculkan data lain. Rancangan *scraper* digunakan untuk memudahkan pada saat proses implementasi ke dalam kode. Dibutuhkan lokasi unik data-data yang akan diambil pada halaman *website*. Penentuan lokasi unik data pada halaman *website* dapat dibantu dengan menggunakan fitur pada *browser*. Secara umum, *flow* alur kerja *data scraper* yang dibangun dapat dilihat pada Gambar 4.1.



Gambar 4.1 Work Flow Data Scraper.

4.1.1 Tokopedia

Pada platform Tokopedia terdapat beberapa informasi yang diperlukan yaitu sub kategori makanan dan minuman, informasi produk, *URL* UMKM, dan informasi UMKM.

A. Sub-Kategori Makanan dan Minuman Tokopedia

Kategori, sub, dan sub-sub kategori pada Tokopedia dapat diakses pada halaman awal, yaitu Tokopedia.com. Terdapat beberapa elemen yang perlu diinteraksi untuk menampilkan informasi kategori. Rancangan interaksi elemen pada halaman awal Tokopedia dapat dilihat pada Tabel 4.1.

Tabel 4.1 Rancangan Interaksi Elemen Halaman Awal Tokopedia.

Elemen	<i>ID Value</i>	Aksi	Deskripsi
div:class	class: css-vk082c	<i>hover</i>	Menampilkan kategori
a:class	class: css-me46ht	<i>hover</i>	Menampilkan sub dan sub-sub kategori

Kategori makanan dan minuman dibagi menjadi beberapa sub dan sub-sub kategori. Tiap sub dan sub-sub kategori memiliki *URL* yang berbeda. Tiap *URL* untuk tiap kategori dibutuhkan untuk diambil dan akan menjadi masukan di proses berikutnya.

Terdapat dua informasi yang dibutuhkan pada bagian kategori, yaitu nama dan *URL* kategori. Selain nama dan *URL* dibutuhkan juga lokasi elemen terhadap dua informasi tersebut dalam struktur halaman. Rancangan data kategori di Tokopedia yang akan disimpan beserta lokasi elemennya dapat dilihat pada Tabel 4.2 dan Tabel 4.3.

Tabel 4.2 Rancangan Data Kategori Tokopedia.

Nama Kolom	Tipe Data	Deskripsi
kategori	varchar(30)	Nama kategori
url_kategori	varchar(100)	<i>URL</i> untuk mengakses halaman kategori produk

Tabel 4.3 Rancangan Lokasi Elemen Data Kategori Tokopedia.

Nama Kolom	Elemen	ID Value	Tipe Elemen
kategori	div:class	1. class: css-1qaqbbz 2. class: css-1nykm5o	<i>text</i>
url_kategori	div:class	1. class: css-1qaqbbz 2. class: css-1nykm5o	<i>href</i>

B. Informasi Produk Makanan dan Minuman Tokopedia

URL kategori digunakan untuk mengakses halaman kategori produk. Halaman kategori produk pada Tokopedia menampilkan semua produk yang sedang dijual pada seluruh kota yang tersedia. Diperlukan adanya perlakuan khusus pada *URL* kategori sehingga dapat mengakses produk yang dijual di Surabaya. Dengan menggunakan fitur *filter* lokasi dan memilih Surabaya, terdapat perubahan pada *URL* kategori. Dari perubahan *URL* tersebut didapatkan bahwa kode kota Surabaya di tokopedia adalah **252**. Kode kota ditambahkan dalam *URL* kategori untuk menampilkan produk yang dijual di Surabaya.

Produk ditampilkan dalam beberapa halaman kategori produk. Perlu adanya interaksi untuk mengganti dari satu halaman ke halaman berikutnya sehingga semua halaman yang menampilkan informasi produk dapat diambil datanya. Rancangan interaksi elemen pada halaman kategori produk Tokopedia dapat dilihat pada Tabel 4.4.

Tabel 4.4 Rancangan Interaksi Elemen Halaman Kategori Produk Tokopedia.

Elemen	ID Value	Aksi	Deskripsi
i:class	class: css-98hn3t e19tp72t1	<i>click</i>	Menampilkan halaman berikutnya pada halaman kategori produk Tokopedia

Pada tiap halaman kategori terdapat dua informasi yang dibutuhkan untuk setiap produk, yaitu nama dan *url* produk. Dibutuhkan lokasi elemen terhadap dua informasi tersebut dalam struktur halaman. Selain informasi nama, dan *URL*, diperlukan juga nama kategori dalam data sehingga dapat diketahui dari kategori apa data produk diambil. Rancangan data produk di Tokopedia yang akan disimpan beserta lokasi elemennya dapat dilihat pada Tabel 4.5 dan Tabel 4.6.

Tabel 4.5 Rancangan Data Produk Tokopedia.

Nama Kolom	Tipe Data	Deskripsi
kategori	varchar(30)	Nama kategori (diambil dari data kategori)
title	varchar(255)	Nama produk
url_produk	varchar(255)	<i>URL</i> untuk mengakses halaman produk

Tabel 4.6 Rancangan Lokasi Elemen Data Produk Tokopedia.

Nama Kolom	Elemen	ID Value	Tipe Elemen
kategori	div:class// div:class	<ul style="list-style-type: none"> • class: css-bk6tzz e1nlzfl3// • class: css-1bjwylw 	<i>text</i>
url_produk	div:class// a	class:css-bk6tzz e1nlzfl3	<i>href</i>

C. URL UMKM Makanan dan Minuman Tokopedia

Terdapat dua jenis *URL* produk pada Tokopedia. *URL* produk dibedakan menjadi *URL* normal dan *URL* promo. Struktur kedua jenis *URL* dapat dilihat pada Tabel 4.7.

Tabel 4.7 Jenis *URL* Produk di Tokopedia.

Jenis URL	Struktur URL
<i>URL</i> normal	https://www.tokopedia.com/nama-UMKM/nama-produk
<i>URL</i> promo	https://ta.tokopedia.com/link-promo?https://www.tokopedia.com/nama-UMKM/nama-produk

Masing-masing struktur *URL* dilakukan *preprocess* sehingga menghasilkan *URL* UMKM. Selain ekstraksi, dilakukan penghapusan duplikasi *URL* untuk menghilangkan redudansi data, sehingga didapatkan *URL* UMKM yang unik. Hasil akhir struktur *URL* UMKM adalah sebagai berikut.

https://www.tokopedia.com/nama-UMKM

Hasil dari *preprocess* akan disimpan menjadi data *URL* UMKM. Rancangan data dan *preprocess URL* UMKM dapat dilihat pada Tabel 4.8.

Tabel 4.8 Rancangan Data dan *Preprocess URL* UMKM Tokopedia.

Nama Kolom	Tipe Data	<i>Preprocess</i>
kategori	varchar(30)	tidak ada <i>preprocess</i>
title_produk	varchar(255)	tidak ada <i>preprocess</i>
url_UMKM	varchar(255)	<ol style="list-style-type: none"> 1. Menghilangkan duplikasi url_UMKM 2. Melakukan <i>split</i> untuk menghilangkan <i>URL</i> produk dan <i>URL</i> promo

D. Informasi UMKM Makanan dan Minuman Tokopedia

Data *URL* UMKM menjadi masukan untuk mengakses halaman UMKM. Setiap *URL* UMKM diakses oleh *scraper* dan menampilkan halaman UMKM. Halaman UMKM dibagi menjadi tiga tampilan, yaitu tampilan awal, info, dan statistik UMKM. Tiap tampilan UMKM akan ditampilkan setelah berinteraksi dengan elemen tertentu. Terdapat elemen yang perlu dilakukan interaksi pada tampilan awal untuk mengakses sebuah informasi. Rancangan interaksi elemen dapat dilihat pada Tabel 4.9.

Tabel 4.9 Rancangan Interaksi Elemen Halaman UMKM Tokopedia.

Elemen	ID Value	Aksi	Deskripsi
div:class	class: css-ypd15i-unf-tooltip	<i>hover</i>	Menampilkan informasi <i>reputation points</i> pada tampilan awal
button:class	class: css-rhf1fq-unf-btn e1ggrouw00	<i>click</i>	Menampilkan tampilan info UMKM
button:class	class: css-1shlf6w-unf-modal__icon e1s17ekyl	<i>click</i>	Menampilkan tampilan statistik UMKM

Data informasi UMKM dibedakan berdasarkan tampilan. Pada tampilan awal terdapat informasi nama, tipe *seller*, kota, jumlah *follower*, *reputation point*, jumlah produk terjual, etalase, *rating*, dan jumlah *review*. Pada tampilan info UMKM terdapat informasi pemilik, deskripsi, alamat, waktu bergabung, jasa pengiriman yang digunakan. Pada tampilan statistik UMKM terdapat informasi jumlah pelanggan puas, jumlah pelanggan netral, jumlah pelanggan tidak puas, kecepatan proses, , jumlah produk yang aktif dijual, jumlah transaksi sukses. Rancangan data UMKM Tokopedia dapat dilihat pada Tabel 4.10. Rancangan lokasi elemen data di halaman UMKM Tokopedia dapat dilihat pada Tabel 4.11.

Tabel 4.10 Rancangan Data UMKM Tokopedia.

Nama Kolom	Tipe Data	Deskripsi
link	varchar (255)	Didapatkan dari data <i>URL</i> UMKM
name	varchar(30)	Nama UMKM
seller_type	varchar(255)	Tipe UMKM
city	varchar(30)	Kota asal UMKM
followers	int	Jumlah <i>followers</i>
reputation_point	int	Jumlah <i>reputation points</i>
product_sold	int	Jumlah produk terjual
rating	int	<i>Rating</i> UMKM
review	int	Jumlah <i>review</i>
numOf_storefront	int	Jumlah etalase
storefront	varchar(255)	Macam-macam etalase yang tersedia pada UMKM
owner	varchar(30)	Pemilik UMKM
description	varchar (255)	Deskripsi UMKM
address	varchar(255)	Alamat UMKM
since	datetime	Bulan dan tahun UMKM

		bergabung di Tokopedia
numOf_delivery_services	int	Jumlah layanan pengiriman yang didukung
delivery_services	varchar(255)	Jasa layanan pengiriman yang didukung
numOf_satisfied	int	Jumlah pelanggan puas
numOf_neutral	int	Jumlah pelanggan netral
numOf_unsatisfied	int	Jumlah pelanggan tidak puas
process_speed	varchar(30)	Kecepatan proses pesanan
active_product	varchar (255)	Jumlah produk yang aktif dijual
success_transaction	varchar(255)	Jumlah transaksi sukses

Tabel 4.11 Rancangan Lokasi Elemen Data UMKM Tokopedia.

Nama Atribut	Elemen	ID Value	Tipe Elemen
link	Didapatkan dari data <i>URL</i> UMKM		
name	div:class/span[index:1]	class: css-1yw0bav	<i>text</i>
seller_type	div:class	class: css-1paptb0-unf-	<i>text</i>

		heading e1qvo2ff6	
city	div:class/li[ind ex=2]/p	class: css- 1k56vr7	<i>text</i>
followers	div:class	class: css- jsut4p-unf- heading e1qvo2ff6	<i>text</i>
reputation_p oint	div:class	class: css- 1ac807k e1hrfe840	<i>text</i>
product_sold	div:class	class: css- lzwncz-unf- heading e1qvo2ff2	<i>text</i>
rating	div:class	class: css- rfs3ih-unf- heading e1qvo2ff2	<i>text</i>
review	div:class	class: css- 1s96mum- unf-heading e1qvo2ff6	<i>text</i>
numOf_store front	div:class	class: css- z2vy7e	<i>count</i>
storefront	div:class	class: css- 17mrx6g	<i>text</i>
owner	div:class	class: css- 1ag3tdd-unf- heading e1qvo2ff4	<i>text</i>

description	div:class	class: css-f4j8r2-unf-heading e1qvo2ff8	<i>text</i>
address	div:class	class: css-f4j8r2-unf-heading e1qvo2ff8	<i>text</i>
since	div:class/p[index:3]	class: css-1c0g9ad e1ufc1ph0	<i>text</i>
numOf_delivery_services	div:class	class: css-2ibdj3	<i>count</i>
delivery_services	div:class	class: css-15n97o5 e1ufc1ph0	<i>text</i>

4.1.2 Shopee

Scraping pada platform shopee akan mengambil beberapa data terkait sub-kategori dari makanan dan minuman, informasi produk, dan informasi UMKM. Halaman awal Shopee adalah shopee.co.id.

A. Sub-Kategori Makanan dan Minuman Shopee

Data terkait kategori dan sub-kategori dari makanan dan minuman pada Shopee dapat diakses pada <https://shopee.co.id/Makanan-Minuman-cat.157>. Perlu adanya interaksi dengan elemen tertentu untuk memunculkan semua sub-kategori dari makanan dan minuman. Rancangan interaksi pada halaman kategori produk Shopee dapat dilihat pada Tabel 4.12.

Tabel 4.12 Rancangan Interaksi Elemen Halaman Kategori Produk Shopee.

Elemen	ID Value	Aksi	Deskripsi
div:class	class: stardust-icon stardust-icon- arrow-down	<i>click</i>	Menampilkan seluruh kategori dan sub-kategori makanan dan minuman

Setelah berinteraksi dengan elemen, sub-kategori makanan dan minuman dimunculkan. Terdapat dua informasi yang dibutuhkan pada bagian kategori, yaitu nama kategori dan sub-kategori beserta *URL*. Selain nama dan *URL* dibutuhkan juga lokasi elemen data terhadap dua informasi tersebut dalam struktur halaman. Rancangan data kategori di Shopee yang akan disimpan dapat dilihat pada Tabel 4.13, sedangkan lokasi elemen datanya dapat dilihat pada Tabel 4.14.

Tabel 4.13 Rancangan Data Kategori Produk Shopee.

Nama Kolom	Tipe Data	Kategori
kategori	varchar(30)	Nama kategori dan sub-kategori makanan dan minuman Shopee
url_kategori	varchar(100)	<i>URL</i> kategori dan sub-kategori makanan dan minuman Shopee

Tabel 4.14 Rancangan Lokasi Elemen Data Kategori Produk Shopee.

Nama Atribut	Elemen	ID Value	Tipe Elemen
kategori	div:class	1. class: shopee-category-list__main-category__link 2. class: shopee-category-list__sub-category	<i>text</i>
url_kategori	div:class	1. class: shopee-category-list__main-category__link 2. class: shopee-category-list__sub-category	<i>href</i>

B. Informasi Produk Makanan dan Minuman Shopee

Halaman kategori produk Shopee diakses menggunakan data *URL* kategori produk. Halaman kategori produk pada Shopee menampilkan semua produk yang sedang dijual pada seluruh area yang tersedia. Diperlukan adanya perlakuan khusus pada *URL* kategori sehingga dapat mengakses produk yang dijual di area yang diinginkan. Pemilihan lokasi dapat dilakukan dengan menggunakan fitur *filter* lokasi. Area Surabaya tidak tersedia pada *filter* lokasi pada Shopee, sehingga untuk mendapatkan data produk yang dijual di Surabaya dipilih lokasi Jawa Timur.

Terjadi perubahan pada *URL* kategori setelah memilih lokasi. Dari perubahan yang terjadi, didapatkan perlunya penambahan ***?locations=Jawa%2520Timur*** pada bagian akhir *URL* kategori produk. Hasil akhir struktur *URL* kategori adalah sebagai berikut.

<https://shopee.co.id/url-kategori/?locations=Jawa%2520Timur>

Produk ditampilkan dalam beberapa halaman kategori produk. Perlu adanya interaksi untuk mengganti dari satu halaman ke halaman berikutnya sehingga semua halaman yang menampilkan informasi produk dapat diambil datanya. Rancangan interaksi elemen pada halaman kategori produk Shopee dapat dilihat pada Tabel 4.15.

Tabel 4.15 Rancangan Interaksi Elemen Halaman Kategori Produk Shopee.

Elemen	ID Value	Aksi	Deskripsi
button:class	class: shopee- button- outline shopee-mini- page- controller__n ext-btn	<i>click</i>	Menampilkan lembar berikutnya pada halaman kategori produk Shopee

Pada halaman kategori produk terdapat dua informasi produk yang dibutuhkan, yaitu nama dan *URL* produk. Dibutuhkan lokasi elemen terhadap dua informasi tersebut dalam struktur halaman. Selain informasi nama, dan *URL*, perlu diketahui dari kategori mana produk diambil sehingga dimasukkan nama kategori yang didapat dari proses pengambilan data sub-kategori. Rancangan data produk di Shopee yang akan disimpan dapat dilihat pada Tabel 4.16. Rancangan lokasi elemen data dapat dilihat pada Tabel 4.17.

Tabel 4.16 Rancangan Data Produk Shopee.

Nama Kolom	Tipe Data	Deskripsi
kategori	varchar(30)	Kategori produk (didapatkan dari data <i>URL</i> kategori)
title	varchar(255)	Nama produk
url_produk	varchar(255)	<i>URL</i> produk

Tabel 4.17 Rancangan Lokasi Elemen Data Produk Shopee.

Nama Atribut	Elemen	ID Value	Tipe Elemen
kategori	Didapatkan dari data <i>URL</i> kategori		
title	div:class	class: O6wiAW	<i>text</i>
url_produk	div:class/ div/a	class: col-xs-2-4 shopee-search- item-result__item	<i>href</i>

C. *URL* UMKM Makanan dan Minuman Shopee

Data *URL* produk digunakan untuk mengakses halaman produk Shopee. Setiap *URL* produk diakses oleh *scraper* untuk mendapatkan *URL* UMKM. Pada halaman produk Shopee, terdapat bagian yang menampilkan informasi mengenai UMKM. Rancangan data *URL* UMKM dapat dilihat pada Tabel 4.18. Rancangan lokasi elemen data *URL* UMKM Shopee dapat dilihat pada Tabel 4.19.

Tabel 4.18 Rancangan Data *URL* UMKM Shopee.

Nama Kolom	Tipe Data	Deskripsi
url_UMKM	varchar(255)	<i>URL</i> UMKM

Tabel 4.19 Rancangan Lokasi Elemen Data Produk Shopee.

Nama Atribut	Elemen	ID Value	Tipe Elemen
url_UMKM	div:class/ a	class: _1jOO4S	<i>href</i>

D. Informasi UMKM Makanan dan Minuman Shopee

Setiap halaman UMKM diakses oleh *scraper* dengan menggunakan masukan data *URL* UMKM. Pada halaman UMKM terdapat beberapa informasi yang dibutuhkan. Rancangan data informasi UMKM Shopee dapat dilihat pada Tabel 4.20. Rancangan lokasi elemen data informasi UMKM Shopee dapat dilihat pada Tabel 4.21.

Tabel 4.20 Rancangan Data UMKM Shopee.

Nama Kolom	Tipe Data	Deskripsi
link	varchar(30)	<i>URL</i> UMKM (didapatkan dari data <i>URL</i> UMKM)
name	varchar(255)	Nama UMKM
city	varchar(255)	Regional UMKM
seller_type	varchar(30)	Tipe seller
product_category	int	Etalase produk
numOf_product	int	Jumlah produk

chat_performance	varchar(30)	Kecepatan membalas pesan
followers	varchar(30)	Pengikut UMKM
rating_review	varchar(30)	Nilai rating dan review UMKM
since	varchar(30)	Waktu bergabung UMKM
voucher	boolean	Terdapat voucher atau tidak
numOf_media	int	Jumlah media yang digunakan
description	varchar(30)	Deskripsi produk

Tabel 4.21 Rancangan Lokasi Elemen Data UMKM Shopee.

Nama Atribut	Elemen	ID Value	Tipe Elemen
link	Didapatkan dari data <i>URL</i> UMKM		
name	div:class	class: section-seller-overview-horizontal__portrait-name"	<i>text</i>
city	div:class	class: _3amru2	<i>text</i>
seller_type	div:class	class: ofs-header__page-name	<i>text</i>
product_category	div:class	class: shopee-category-list__body	<i>text</i>

numOf_product	div:class/ div[index: 2]	<ul style="list-style-type: none"> • class: section-seller-overview__item-text • <i>contains</i>('produk') 	<i>text</i>
chat_performance	div:class/ div[index: 2]	<ul style="list-style-type: none"> • class: section-seller-overview__item-text • <i>contains</i>('Performa chat') 	<i>text</i>
followers	div:class/ div[index: 2]	<ul style="list-style-type: none"> • class: section-seller-overview__item-text • <i>contains</i>('pengikut') 	<i>text</i>
rating_review	div:class/ div[index: 2]	<ul style="list-style-type: none"> • class: section-seller-overview__item-text • <i>contains</i>('penilaian') 	<i>text</i>
since	div:class/ div[index: 2]	<ul style="list-style-type: none"> • class: section-seller-overview__item-text • <i>contains</i>('bergabung') 	<i>text</i>
voucher	div:class	class: shop-page__section	<i>boolean</i>

		shop- page__vouchers	
numOf_med ia	div:class/ div	class: image- carousel__dots	<i>count</i>
description	div:class	class:contains('sh op-page-shop- description')	<i>text</i>

4.1.3 Google My Business

Terdapat beberapa tahap dalam proses *scraping* di platform Google My Business. Dibutuhkan beberapa rancangan yang mendukung proses *scraping* Google My Business. Rancangan tersebut diantaranya adalah rancangan data pendukung, dan informasi UMKM.

A. Data Pendukung *Scraping* Google My Business

Dalam melakukan *scraping* Google My Business, diperlukan adanya *keyword* atau kata kunci pencarian yang menjadi masukan dalam mencari UMKM. Kata kunci pencarian tersusun dari dua hal, yaitu kategori dan lokasi.

Perlu adanya beberapa data pendukung dalam penyusunan kata kunci pencarian. Data pendukung yang dibutuhkan adalah data kategori Google My Business, dan data kecamatan dan kelurahan di Surabaya sebagai lokasi.

1) Data Kategori Google My Business

Kategori Google My Business menjadi penyusun kata kunci pencarian pada proses berikutnya. Pencarian kategori dilakukan secara manual menelusuri internet. Kategori dipilih berdasarkan relevansi dengan makanan dan minuman. Kategori yang relevan dipilih dan disimpan untuk digunakan pada proses berikutnya. Rancangan data kategori Google My Business dapat dilihat pada Tabel 4.22.

Tabel 4.22 Rancangan Data Kategori Google My Business.

Nama Kolom	Tipe Data	Deskripsi
kategori	varchar(100)	Kategori Google My Business yang relevan dengan makanan dan minuman

2) Data Kecamatan dan Kelurahan

Data kecamatan dan kelurahan dibutuhkan untuk digunakan sebagai penyusun kata kunci pencarian pada proses berikutnya. Data diambil dari *website* Pemerintah Kota Surabaya [16]. Rancangan data kecamatan dan kelurahan untuk *scraping* Google My Business dapat dilihat pada Tabel 4.23.

Tabel 4.23 Rancangan Data Kecamatan dan Kelurahan.

Nama Kolom	Tipe Data	Deksripsi
kecamatan	varchar(100)	Nama kecamatan di Surabaya
kelurahan	varchar(100)	Nama kelurahan di Surabaya

B. Informasi UMKM Google My Business

Pada tahap ini dilakukan *scraping* data untuk mendapatkan informasi UMKM Google My Business yang relevan dengan kategori makanan dan minuman di Surabaya. Data pendukung, yaitu data kategori, serta kecamatan dan kelurahan, digunakan sebagai masukan dalam penyusunan kata kunci pencarian. Diperlukan interaksi dengan beberapa elemen dalam tahap

scraping informasi UMKM. Rancangan interaksi elemen dapat dilihat pada Tabel 4.24.

Tabel 4.24 Rancangan Interaksi Google My Business.

Elemen	ID Value	Aksi	Deskripsi
div:class	class: gLFyf gsfi	<i>fill field</i>	Memasukkan kata kunci pencarian dalam <i>search field</i>
div:class	class: zkIadb	<i>click</i>	Masuk ke halaman Google Maps My Business
div:class//div:class	1. class: VkpG Bb 2. class: C8TU Kc rllt_1 ink a-no-hover-decoration	<i>click</i>	Menampilkan informasi tiap UMKM yang ditemukan
div:id	id: pnnext	<i>click</i>	Menuju halaman hasil pencarian

			berikutnya
div: class/div/div/div[index:1]/div[index:1]	class: zloOqf PZPZlf	click	Menampilkan informasi jam operasional UMKM

Google My Business memberikan beberapa informasi UMKM dari hasil pencarian yang akan diambil dan disimpan menjadi satu data, yaitu data UMKM Google My Business. Rancangan data UMKM Google My Business dapat dilihat pada Tabel 4.25. Informasi-informasi tersebut tersedia dalam elemen-elemen tertentu dalam struktur halaman *website*. Rancangan lokasi elemen data UMKM Google My Business dapat dilihat pada Tabel 4.26.

Tabel 4.25 Rancangan Data UMKM Google My Business.

Nama Kolom	Tipe Data	Deskripsi
name	varchar(100)	Nama UMKM
website	varchar(100)	Website milik UMKM
rating	float	Rating UMKM
numOf_google_review	varchar(100)	Jumlah review
price	varchar(100)	Kategori harga produk yang

		dijual oleh UMKM
category	varchar(100)	Kategori UMKM
address	varchar(100)	Alamat UMKM
operational_hours	varchar(100)	Jam operasional UMKM
phone	int	Nomor telpon UMKM
instagram	varchar(100)	Akun Instagram UMKM

Tabel 4.26 Rancangan Lokasi Elemen Data UMKM Google My Business.

Nama Atribut	Elemen	ID Value	Tipe Elemen
name	div:class/div/span	class: SPZz6b	<i>text</i>
website	div:class/a	class: QqG1Sd	<i>href</i>
rating	div:class/div/span[index:1]	class: Ob2kfd	<i>text</i>
numOf_google_review	div:class/span/a/span	class: hqzQac	<i>text</i>

price	div:class[index:1]	class: YhemCb	<i>text</i>
category	div:class[index:2]	class: YhemCb	<i>text</i>
address	div:class	class: LrzXr	<i>text</i>
operational_ hours	div:class[index:1]/ /div/div/div[index:1] /div[index:2]/table/t body	class: zloQf PZPZlf	<i>text</i>
phone	div:class	class: LrzXr zdqRlf kno-fv	<i>text</i>
instagram	div:class/ div/g- link/a	class: PZPZlf kno-vrt-t	<i>href</i>

4.1.4 Instagram

Scraping pada Instagram menggunakan masukan dari nama UMKM Tokopedia dan Shopee. Nama UMKM menjadi kata kunci pencarian pada *search field* Instagram sehingga ditampilkan hasil yang relevan dengan nama UMKM Tokopedia dan Shopee. Diperlukan akses *login* menggunakan sebuah akun agar dapat melakukan pencarian di halaman Instagram. Maka, dibutuhkan akun untuk melakukan *scraping*. Beberapa interaksi elemen dalam halaman Instagram dapat dilihat pada Tabel 4.27. Data yang dibutuhkan adalah data *account*, deskripsi akun, dan *URL* akun. Rancangan data yang akan disimpan pada proses *scraping* Instagram dapat dilihat pada Tabel 4.29.

Tabel 4.27 Rancangan Interaksi Elemen Pada Instagram.

Elemen	ID Value	Aksi	Deskripsi
div:name	name: username	<i>send_keys</i>	Memasukkan <i>username</i> akun Instagram ke dalam <i>field username</i>
div:name	name: password	<i>send_keys</i>	Memasukkan <i>password</i> akun Instagram ke dalam <i>field password</i>
div:class	class: aOOIW HoLwm	<i>click</i>	Mengklik untuk menutup sebuah <i>popup</i> untuk mencari teman
div:class/ input	class: LWmhU _0aCwM	<i>send_keys</i>	Memasukkan nama UMKM pada kolom pencarian

Tabel 4.28 Rancangan Data Hasil Pencarian Pada Instagram.

Nama Kolom	Tipe Data	Deksripsi
account	varchar(255)	Nama akun instagram
description	varchar(255)	Deskripsi pada akun instagram
link	varchar(255)	<i>URL</i> akun instagram

Tabel 4.29 Rancangan Lokasi Elemen Data Hasil Pencarian Pada Instagram.

Atribut	Elemen	ID Value	Tipe Elemen
Account	div:class	class="Ap253"	<i>text</i>
Description	div:class	class="Fy4o8"	<i>text</i>
Link	div:class	class="Fy4o8"	<i>href</i>

4.2 Rancangan Data Integration

Seluruh data informasi UMKM dari ketiga platform akan diintegrasikan dengan metode *Entity Resolution*. *Entity Resolution* dijalankan menggunakan teknik *similarity*. Penerapan metode *entity resolution* dilakukan dengan melihat nilai atau tipe atribut data yang sama. Atribut data yang dipilih antara lain *link*, *phone*, *website*, *instagram* dan *name* karena masing-masing platform memiliki data tersebut. Library *python* yaitu *SequenceMatcher* akan digunakan dalam penerapan tes *similarity*.

Dibutuhkan *threshold* atau ambang batas minimum sehingga hasil *similarity* dianggap sebagai *match* atau UMKM yang sama. Similaritas atribut data akan diterapkan prioritas secaraurut dari yang tinggi hingga yang paling rendah. Urutanprioritasnya adalah *link*, *phone*, *website*, *instagram* dan *name*. Setiap atribut memiliki nilai *threshold* yang berbeda disesuaikan dengan kondisi data. Nilai *threshold* ditentukan dengan banyaknya nilai benar atau *True Positive (1.00)* dibandingkan dengan jumlah keseluruhan hasil *match* atau UMKM yang sama. Hasil *similarity* yang bernilai sama atau lebih besar dari nilai *threshold* akan dianggap sama atau *match*.

Tabel 4.30 Rancangan *Data Integration*.

Urutan	Sumber Data	Data yang akan Diintegrasikan	Keterangan
1	Tokopedia	Instagram_Tokopedia	UMKM daring pada Tokopedia dengan tambahan informasi mengenai platform Instagram
2	Shopee	Instagram_Shopee	UMKM daring pada Shopee dengan tambahan informasi mengenai platform Instagram
3	Tokopedia	Shopee	UMKM daring secara umum
4	Google Business	Tokopedia dan Shopee	UMKM secara keseluruhan (daring dan tidak daring)

4.3 Perhitungan Kualitas Data Integrasi

Dalam tahap perhitungan kualitas data integrasi, sebagian besar merupakan hasil saduran dari hasil pekerjaan tugas akhir Syafrie Dwi Faisal [17]. Kualitas data diukur dengan menggunakan tingkat presisi atau *precision* dari hasil *matching* pada tahap integrasi data. Perhitungan presisi dimulai dari menentukan *sample* yang akan diuji dari hasil integrasi, melakukan pengujian, dan analisis hasil pengujian. Pengujian kualitas data integrasi dilakukan pada ketiga proses integrasi, yaitu integrasi data Tokopedia dengan Shopee, Google My Business dengan Tokopedia, serta Google My Business dengan Shopee.

4.3.1 Menentukan *Confidence Level*

Confidence level (α) menunjukkan seberapa yakin dengan hasil integrasi bahwa hasil tersebut akan sama apabila dilakukan pengulangan. Dalam penelitian ini digunakan *confidence level* sebesar 90%.

4.3.2 Perhitungan Jumlah *Sample*

Perhitungan jumlah *sample* dilakukan dengan menggunakan metode *Krejcie and Morgan*. Rumus dari metode *Krejcie and Morgan* dapat dilihat sebagai berikut [18].

$$s = \frac{X^2 NP(1 - P)}{d^2(N - 1) + X^2 P(1 - P)}$$

Dimana s merupakan jumlah *sample* yang ingin ditentukan. X^2 merupakan nilai *chi* kuadrat ($df=1$) yang dapat dilihat pada . N merupakan jumlah populasi, di mana dalam penelitian ini populasi yang digunakan adalah hasil integrasi yang dianggap *match*. P merupakan proporsi populasi (diasumsikan 0,5 karena akan memberikan jumlah *sample* yang maksimum). d merupakan derajat tingkat error (0,05). Dengan menggunakan metode *Krejcie and Morgan*, *confidence level* yang digunakan adalah 95%.

4.3.3 Pengambilan *Sample*

Sample diambil menggunakan metode *simple random sampling*. *Sample* diambil secara acak pada data hasil integrasi. *Sample* yang digunakan diambil dari populasi jumlah hasil *match* pada tahap integrasi.

4.3.4 Validasi data

Validasi data dilakukan secara manual. Tiap data UMKM yang dianggap memiliki *match* dari hasil integrasi dilakukan pengecekan manual apakah UMKM yang dianggap sama tersebut benar-benar sama atau tidak. Hasil yang dianggap sama dan benar-benar sama disebut sebagai *True Positive* (TP), sedangkan hasil yang dianggap sama tetapi ternyata UMKM yang berbeda disebut dengan *False Positive* (FP).

4.3.5 Perhitungan Nilai Presisi

Kualitas data dilihat dari perhitungan nilai presisi yang didapatkan dari hasil integrasi. Nilai presisi didapatkan dengan menghitung rasio hasil *match* benar disebut dengan *True Positive* (TP) dibandingkan dengan keseluruhan hasil yang dianggap *match* atau hasil *match* benar (TP) ditambah hasil *match* salah atau disebut dengan *False Negative* (FP). Perhitungan nilai presisi dapat dilihat sebagai berikut.

$$\text{Presisi} = \frac{TP}{TP + FP}$$

4.4 Rancangan Visualisasi dan Analisis

Rancangan visualisasi dan matriks yang digunakan merupakan hasil saduran dari pekerjaan tugas akhir Aldy Syah Daviq Ramadhan [19]. Pada proses visualisasi dan analisis, hasil data yang telah dikumpulkan dan diolah menjadi bahan masukan. *Software* yang digunakan adalah Microsoft PowerBI. Diperlukan dua komponen penyusun visualisasi, yaitu matriks, dan diagram.

4.4.1 Matriks

Berdasarkan atribut-atribut yang telah dirancang, perlu adanya matriks untuk mengukur dan menganalisis data yang dikumpulkan. Matriks menjadi acuan dalam melakukan visualisasi. Rancangan matriks dapat dilihat pada Tabel 4.31.

Tabel 4.31 Rancangan Matriks Visualisasi.

Platform	Atribut	Metrics	Tujuan
Offline	Index	Jumlah UMKM	Melihat jumlah UMKM pada perspektif tertentu
	Review	Rata-rata review	Melihat bagaimana nilai tengah dari jumlah review pada perspektif tertentu
	Operational_hours	Rata-rata jumlah jam kerja	Melihat intensitas aktivitas kerja UMKM pada perspektif tertentu
	Rating	Rata-rata rating	Melihat bagaimana nilai tengah dari rating pada perspektif tertentu
	Rating, Operational_hours	Korelasi rating dengan Jam kerja	Melihat bagaimana hubungan antara variabel rating dengan jam kerja pada perspektif tertentu
Online	Index	Jumlah UMKM	Melihat jumlah UMKM pada perspektif tertentu
	Active_product/nu	Rata-rata jumlah	Melihat nilai tengah dari jumlah produk yang aktif dijual

	mOf_product	katalog produk	pada UMKM pada perspektif tertentu
	Follower	Rata-rata jumlah follower	Melihat nilai tengah dari jumlah follower pada perspektif tertentu
	Review	Rata-rata review	Melihat bagaimana nilai tengah dari jumlah review pada perspektif tertentu
	Rating	Rata-rata rating	Melihat bagaimana nilai tengah dari rating pada perspektif tertentu
	Seller_type	Jumlah UMKM berdasarkan jenis UMKM	Melihat bagaimana persebaran jenis UMKM pada perspektif tertentu
	Rating, product_sold	Korelasi rating dengan jumlah produk yang dijual	Melihat bagaimana hubungan antara rating dengan jumlah produk yang dijual oleh UMKM
	Delivery_services	Mode dari layanan pengiriman	Mengetahui frekuensi terbesar dari layanan pengiriman yang digunakan
	Product_category	Mode dari kategori produk	Mengetahui frekuensi terbesar dari kategori produk yang ada

Online Offline	Index	Jumlah UMKM	Melihat jumlah UMKM pada perspektif tertentu
	Review	Rata-rata review	Melihat bagaimana nilai tengah dari jumlah review pada perspektif tertentu

4.4.2 Visualisasi Data

Rancangan visualisasi informasi dari hasil *scraping* disandur dari pekerjaan tugas akhir Aldy Syah Daviq Ramadhan [19]. Visualisasi data dapat dilakukan dengan menggunakan berbagai media, salah satunya adalah diagram. Perancangan visualisasi yang akan digunakan untuk menggambarkan data sesuai dengan rancangan matriks yang dibuat. Rancangan visualisasi dapat dilihat pada Tabel 4.32.

Tabel 4.32 Rancangan Visualisasi Data.

Platform	Metrics	Jenis Visualisasi
Offline	Jumlah UMKM	<i>Card, bar chart, map</i>
	Rata-rata review	<i>Card</i>
	Rata-rata jumlah jam kerja	<i>Card</i>
	Rata-rata rating	<i>Gauge</i>
	Korelasi rating dengan Jam kerja	<i>Scatter chart</i>
Online	Jumlah UMKM	<i>Card, map</i>
	Rata-rata jumlah katalog produk	<i>card</i>
	Rata-rata jumlah follower	<i>card</i>
	Rata-rata review	<i>card</i>
	Rata-rata rating	<i>Gauge</i>

	Jumlah UMKM berdasarkan jenis UMKM	<i>Donut chart</i>
	Korelasi rating dengan jumlah produk yang dijual	<i>Bar Chart</i>
	Mode dari layanan pengiriman	<i>Wordcloud</i>
	Mode dari kategori produk	<i>Wordcloud</i>
Online Offline	Jumlah UMKM	<i>Card</i>
	Rata-rata review	<i>Pie Chart</i>

Halaman ini sengaja dikosongkan.

BAB V IMPLEMENTASI

Pada bab ini dijelaskan implementasi dari rancangan ke dalam bentuk program. Program disusun dari beberapa proses mulai dari *scraping*, *preprocess*, dan visualisasi data.

5.1 *Scraping Data*

Scraping data atau pengumpulan data dari *website* sebagai proses awal untuk menghasilkan kumpulan data dari beberapa kanal daring. *Scraping data* diotomasi menggunakan program. Secara keseluruhan, program dijalankan pada dua perangkat berbeda. Perangkat pertama memiliki *Operating System* (OS) Windows, dan yang kedua merupakan *Virtual Machine* (VM) dengan OS Linux. Pembuatan program menggunakan beberapa *tools* dengan detail sebagai berikut.

- Bahasa pemrograman yang digunakan adalah Python.
- *Library* yang digunakan adalah Selenium, Pandas, Pyvirtualdisplay, dan Time. Selenium sebagai *browser automation and simulator* digunakan untuk menjalankan mesin pencari dengan nama *driver* dan mensimulasikan interaksi dengan elemen, seperti *click* dan *hover*. Pandas untuk menyimpan hasil akhir pengumpulan data ke dalam bentuk *file*. Pyvirtualdisplay sebagai *display simulator* pada VM Linux sehingga program dapat berjalan secara *headless*. Time sebagai *script running control* untuk mengatur jalannya program berdasarkan waktu yang ditentukan, seperti *delay*, *sleep*, dan *wait*.
- Mesin pencari yang digunakan adalah *chromedriver*.
- Metode *scraping* yang digunakan adalah *headless* pada perangkat Linux dan tidak pada perangkat Windows.
- Bahasa kueri yang digunakan untuk menemukan lokasi elemen data adalah XPath.

Struktur program dibagi menjadi tiga, yaitu *load library*, *load driver*, dan *parser*. Pengaturan pada *driver* dilakukan beberapa

hal, yaitu *incognito*, dan tanpa gambar sehingga akses lebih cepat dan ringan. Implementasi dari penggunaan *library*, *load*, dan setting *driver* dapat dilihat sebagai berikut.

```
from selenium import webdriver
import pandas as pd

#ChromeDriver
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument("--incognito")
prefs = {
    "profile.managed_default_content_settings.images": 2}
chrome_options.add_experimental_option("prefs", prefs)
driver = webdriver.Chrome(options=chrome_options)
driver.maximize_window()
```

Scraping dibagi berdasarkan platform, yaitu Tokopedia, Shopee, dan Google My Business. Secara umum dari ketiga platform tersebut, struktur program *scraping* terdiri dari *load library*, *load* file pendukung dan *file* masukan, pengambilan, serta penyimpanan data. Pada bagian pengambilan data terdapat konversi menjadi *syntax* XPath untuk mengakses lokasi elemen data. Penentuan lokasi data menggunakan XPath dapat dibantu dengan menggunakan fitur *XPATH* pada *browser* yang digunakan untuk observasi halaman *website*. *Syntax* XPath yang digunakan adalah sebagai berikut.

```
element_data = driver.find_element_by_xpath('XPath')
data = element_data.tipe_element
```

5.1.1 Tokopedia

Dari hasil rancangan *data scraper* pada bagian 4.1.1, diperlukan adanya konversi sehingga dapat dilakukan pada *tools* yang digunakan. Tahap *scraping* pada Tokopedia dibagi menjadi pengambilan data kategori, produk, dan UMKM yang relevan dengan makanan dan minuman.

A. *Scraping* Informasi Kategori Makanan dan Minuman Tokopedia

Sesuai dengan rancangan *scraping* data kategori Tokopedia, dilakukan beberapa konversi sesuai dengan *tools* yang digunakan. Konversi dilakukan terhadap interaksi elemen, pengambilan dan penyimpanan data. Dalam menemukan lokasi elemen digunakan XPath, dan melakukan aksi pada elemen menggunakan Selenium. Tampilan kategori dan sub-kategori dimunculkan setelah berinteraksi dengan elemen sesuai pada rancangan di bagian 4.1.1A. Konversi rancangan ke dalam implementasi dapat dilihat pada Tabel 5.1. Data kategori diambil sesuai dengan rancangan yang dibuat. Implementasi interaksi elemen dapat dilihat sebagai berikut.

```
wait_kategori_button = WebDriverWait(driver, 3)
.until(EC.visibility_of_element_located(
    (By.XPATH, '//*[@class="css-vk082c"]'))))

ActionChains(driver).move_to_element(
    wait_kategori_button).perform()
```

Dengan struktur umum *scraping* yang dibuat, maka program *scraping* informasi kategori pada Tokopedia dapat dilihat pada bagian LAMPIRAN Tabel A.1.

Tabel 5.1 Implementasi XPath Elemen Lokasi Data Kategori Tokopedia.

Nama Kolom	XPath	Tipe Elemen
kategori	1. <code>//*[@class="css-me46ht"]</code> 2. <code>//*[@class="css-1qaqbbz"]</code>	<i>text</i>
url_kategori	1. <code>//*[@class="css-me46ht"]</code> 2. <code>//*[@class="css-1qaqbbz"]</code>	<i>href</i>

B. *Scraping* Informasi Produk Makanan dan Minuman Tokopedia

Proses *scraping* informasi produk di Tokopedia menggunakan data kategori. *URL* dari data kategori digunakan sebagai alamat tujuan untuk diakses *scraper*. Tiap *URL* akan menjadi alamat tujuan *scraper*. Pada halaman kategori produk, semua produk dari seluruh daerah ditampilkan. Dari rancangan *scraping* informasi produk didapatkan bahwa Surabaya memiliki kode unik, yaitu **252**. Diperlukan penambahan untuk mengubah struktur *URL* kategori pada bagian belakang dengan **?page=1&fcity=252**. Struktur akhir *URL* kategori saat diakses oleh *scraper* adalah sebagai berikut.

<https://www.tokopedia.com/url-kategori?page=1&fcity=252>

Jumlah kategori makanan dan minuman di Tokopedia berjumlah 147, maka *scraper* juga perlu mengakses satu persatu *URL* kategori tersebut. Perulangan mengakses *URL* kategori menggunakan metode *for* pada Python. Pada pembuatan *scraper*, penambahan *URL* dimasukkan saat akan mengakses *URL* kategori dalam bagian kode *for*. Kode perulangan dan akses *URL* kategori dengan penambahan *filter* kota Surabaya dapat dilihat sebagai berikut.

```
for url in list_url_kategori:
    driver.get(str(url) + '?page=1&fcity=252')
```

Pada halaman kategori produk, terdapat beberapa interaksi yang perlu dilakukan sesuai dengan rancangan. Produk ditampilkan dalam beberapa halaman-halaman yang berbeda. Maka, *scraper* melakukan perulangan dengan melakukan interaksi pada tombol *next_page* hingga tombol tersebut tidak dapat diinteraksikan lagi. Perulangan hingga jumlah yang tidak pasti dilakukan menggunakan kode *while* pada Python. Kode perulangan untuk membuka halaman kategori produk berikutnya dapat dilihat sebagai berikut.

```

while True:
    next_page = driver.find_elements_by_xpath
    ('//i[@class="css-98hn3t e19tp72t1"]')

    (get data)

    if(next_page):
        next_page[0].click()

```

Pada saat mengakses setiap halaman kategori produk terdapat informasi yang dikumpulkan. Pengambilan data dilakukan dengan melakukan konvseri rancangan menjadi XPath. Implementasi XPath untuk lokasi elemen data produk dilihat pada tabel Tabel 5.2. Hasil akhir kode program *scraping* informasi produk makanan dan minuman Tokopedia dapat dilihat pada LAMPIRAN Tabel A.2.

Hasil *scraping* informasi produk didapatkan data sejumlah **254.863** baris. Data *URL* produk yang didapatkan masih belum memiliki struktur yang sama sehingga perlu dilakukan *preprocess* seperti yang telah dirancang.

Tabel 5.2 Implementasi XPath Lokasi Elemen Data Produk Tokopedia.

Nama Kolom	XPath	Tipe Elemen
kategori	<code>//*[@class="css-bk6tzz e1nlzf13"]//*[@class="css-1bjwylw"]</code>	<i>text</i>
url_produk	<code>//*[@class="css-bk6tzz e1nlzf13"]/a</code>	<i>href</i>

C. *Preprocess URL* UMKM Makanan dan Minuman Tokopedia

Preprocess URL produk menjadi *URL* UMKM dibagi menjadi beberapa tahap. Tahap pertama adalah memisahkan *URL* promo

dan *URL* normal pada kumpulan data. Pemisahan dilakukan dengan menggunakan ~ (negasi) *contains* sehingga mengambil semua *URL* yang tidak mengandung kata *ta.tokopedia.com/promo* untuk *URL* normal. Pemisahan data pada kode dapat dilihat sebagai berikut.

```
df_url_produk_notpromo = df_url_produk
    [~df_url_produk.link.str.contains(
        "ta.tokopedia.com/promo")]
```

Tahap kedua adalah merubah struktur *URL* promo dan normal menjadi sama dengan struktur yang telah dirancang pada bagian 4.1.1C. Pengubahan struktur *URL* normal adalah dengan melakukan *split* pada *URL* sehingga didapatkan sesuai dengan struktur *URL* UMKM. Pengubahan struktur *URL* promo dilakukan dengan cara menemukan *URL* UMKM yang diawali dengan *https%3A%2F%2F* dan diakhiri dengan *%3Fsrc%3Dtopads*, kemudian mengubahnya menjadi struktur *https://www.tokopedia.com/nama-UMKM/nama-produk*. Pada kode, mengubah struktur dilakukan dengan menghapus *%3A%2F*, mengubah *https%2F* menjadi *https://*, menghapus *%2F*, dan melakukan *split*. Implementasi *preprocess URL* produk promo produk menjadi *URL* UMKM pada kode sebagai berikut.

```
list_toko_promo = []
for url in list_url_produk_promo:
    url = url[url.find(
        'https%3A%2F%2F'):
        url.find('%3Fsrc%3Dtopads')]
    url = url.replace('%3A%2F', '')
    url = url.replace('https%2F', 'https://')
    url = url.replace('%2F', '/')
    list_toko_promo.append(url.rsplit("/",1)[-2])
```

Setelah *URL* promo memiliki struktur sama dengan *URL* normal, maka dilakukan *split* seperti yang dilakukan terhadap *URL* normal. *Preprocess* penghapusan *URL* promo dan *split URL* pada kode dapat dilihat sebagai berikut.

Setelah semua *URL* produk memiliki struktur *URL* UMKM, yaitu *https://www.tokopedia.com/nama-UMKM*, maka dilakukan penghapusan duplikasi *URL*. Duplikasi *URL* dilakukan karena setiap UMKM di Tokopedia memiliki *URL* yang unik, sehingga tidak perlu melakukan *scraping* terhadap *URL* yang sama berulang kali. Penghapusan duplikasi data pada kdoes dapat dilihat sebagai berikut.

```
df_url_toko_all_unique = df_url_toko_all.  
    drop_duplicates('link')
```

Preprocess URL produk menjadi *URL* UMKM menghasilkan kumpulan data sebanyak **23.429** *URL* UMKM yang unik. Implementasi dalam bentuk kode untuk *preprocess URL* produk menjadi *URL* UMKM dapat dilihat pada LAMPIRAN Tabel A.3.

D. *Scraping* Informasi UMKM Makanan dan Minuman Tokopedia

Scraper informasi UMKM menggunakan data *URL* UMKM untuk masukan sebagai alamat akses. Setiap *URL* UMKM diakses oleh *scraper* untuk mengakses halaman informasi UMKM di Tokopedia. Pengulangan akses untuk setiap *URL* UMKM Tokopedia menggunakan kode *for* pada Python dapat dilihat sebagai berikut.

```
for url in list_url_toko:  
    driver.get(url)
```

Saat *scraper* mengakses setiap *URL* UMKM, data-data yang telah dirancang untuk dikumpulkan sesuai dengan rancangan pada bagian 4.1.1D diambil. Terdapat satu interaksi, yaitu interaksi untuk memunculkan *reputation point* pada laman UMKM. Implementasi interaksi *reputation point* dapat dilihat sebagai berikut.

```
#get reputation points
point_hover_element = driver.find_element_by_xpath(
    '//*[@class="css-ypd15i-unf-tooltip"]')
ActionChains(driver).move_to_element(
    point_hover_element).perform()
```

Dalam melakukan pengambilan data, dilakukan konversi rancangan menjadi XPath untuk menemukan lokasi elemen data. Konversi elemen data informasi UMKM menjadi XPath dapat dilihat pada Tabel 5.3.

Setelah mengambil data pada laman UMKM, *scraper* berinteraksi dengan tombol Info UMKM untuk mengakses informasi detail UMKM serta mengambil data-data yang dirancang. Tampilan berikutnya yang dirancang untuk diambil adalah tampilan statistik UMKM. Implementasi interaksi halaman UMKM pada kode dapat dilihat sebagai berikut.

```
#open info toko
info_buttons = driver.find_elements_by_xpath(
    '//*[@class="css-rhf1fq-unf-btn e1ggruw00"]')
driver.execute_script("arguments[0].click();",
    info_buttons[0])

#open statistics
statistic_element = driver.find_elements_by_xpath(
    '//*[@class="css-16z1w0p-unf-heading e1qvo2ff6"]')
driver.execute_script("arguments[0].click();",
    statistic_element[0])
```

Secara keseluruhan, kode *scraping* informasi UMKM dapat dilihat pada LAMPIRAN Tabel A.4.

Tabel 5.3 Implementasi XPath Lokasi Elemen Data UMKM Tokopedia.

Nama Atribut	XPath	Tipe Elemen
link	Didapatkan dari data URL UMKM	

name	//*[@class="css-1yw0bav"]/span[1]	<i>text</i>
seller_type	//*[@class="css-1paptb0-unf-heading e1qvo2ff6"]	<i>text</i>
city	//*[@class="css-1k56vr7"]/li[2]/p	<i>text</i>
followers	//*[@class="css-jsut4p-unf-heading e1qvo2ff6"]	<i>text</i>
reputation_point	//*[@class="css-1ac807k e1hrfe840"]	<i>text</i>
product_sold	//*[@class="css-lzwncz-unf-heading e1qvo2ff2"]	<i>text</i>
rating	//*[@class="css-rfs3ih-unf-heading e1qvo2ff2"]	<i>text</i>
review	//*[@class="css-1s96mum-unf-heading e1qvo2ff6"]	<i>text</i>
numOf_storefront	//*[@class="css-z2vy7e"]	<i>count</i>
storefront	//*[@class="css-17mrx6g"]	<i>text</i>
owner	//*[@class="css-1ag3tdd-unf-heading e1qvo2ff4"]	<i>text</i>
description	//*[@class="css-f4j8r2-unf-heading e1qvo2ff8"]	<i>text</i>
address	//*[@class="css-f4j8r2-unf-heading e1qvo2ff8"]][1]	<i>text</i>
since	//*[@class="css-1c0g9ad e1ufc1ph0"]/p[3]	<i>text</i>
numOf_delivery_services	//*[@class="css-2ibdj3"]	<i>count</i>

delivery_services	//*[@class="css-15n97o5e1ufc1ph0"]	text
-------------------	------------------------------------	------

5.1.2 Shopee

Pengambilan data pada platform shopee dibagi menjadi beberapa tahap, yaitu pengambilan data sub-kategori dari makanan dan minuman, informasi produk, dan informasi UMKM.

A. *Scraping* Sub-Kategori Makanan dan Minuman Shopee

Scraper Data terkait kategori dan sub-kategori dari makanan dan minuman pada Shopee dapat diakses pada <https://shopee.co.id/Makanan-Minuman-cat.157>. Pertama-tama *scraper* perlu untuk menuju *URL* yang menyediakan informasi kategori makanan dan minuman pada Shopee. Implementasi akses halaman kategori makanan dan minuman pada Shopee dapat dilihat sebagai berikut.

```
driver.get(
    'https://shopee.co.id/Makanan-Minuman-cat.157/')
```

Pada halaman kategori makanan dan minuman di Shopee, tidak semua sub-kategori ditampilkan sehingga perlu melakukan interaksi pada elemen tombol *arrow_down_button* untuk memunculkan semua sub-kategori. Implementasi pada kode dapat dilihat sebagai berikut.

```
arrow_down_button = driver.find_elements_by_xpath(
    '//*[@class="stardust-icon stardust-icon-arrow-down"]')
arrow_down_button[0].click()
```

Setelah berinteraksi dengan elemen, sub-kategori makanan dan minuman dimunculkan. *Scraper* menemukan lokasi elemen data dengan menggunakan XPath sesuai dengan rancangan. Terdapat dua informasi yang diambil pada bagian kategori, yaitu nama kategori dan sub-kategori beserta *URL*. Konversi rancangan menjadi XPath dapat dilihat pada Tabel 5.4.

Implementasi kode untuk *scraping* data kategori makanan dan minuman di Shopee dapat dilihat pada LAMPIRAN Tabel A.5.

Tabel 5.4 Implementasi XPath Lokasi Elemen Data Kategori Produk Shopee.

Nama Atribut	ID Value	Tipe Elemen
kategori	1. <code>//*[@class="shopee-category-list__main-category__link"]</code> 2. <code>//*[@class="shopee-category-list__sub-category"]</code>	<i>text</i>
url_kategori	1. <code>//*[@class="shopee-category-list__main-category__link"]</code> 2. <code>//*[@class="shopee-category-list__sub-category"]</code>	<i>href</i>

B. *Scraping* Informasi Produk Makanan dan Minuman Shopee

Halaman kategori produk Shopee diakses menggunakan data *URL* kategori produk. Halaman kategori produk pada Shopee menampilkan semua produk yang sedang dijual pada seluruh area yang tersedia. Pemilihan lokasi dapat dilakukan dengan menggunakan fitur *filter* lokasi. *Filter* lokasi yang digunakan adalah Jawa Timur karena Surabaya tidak tersedia. Didapatkan hasil akhir struktur *URL* kategori sebagai berikut.

`https://shopee.co.id/url-kategori/?locations=Jawa%2520Timur`

Scraper mengakses setiap *URL* kategori produk Shopee dengan menambahkan `?locations=Jawa%2520Timur` ke dalam *URL*.

Implementasi pada kode penambahan *URL* kategori produk Shopee dapat dilihat sebagai berikut.

```
for url in list_url_kategori:
    driver.get(str(url) + '?locations=Jawa%2520Timur')
```

Produk ditampilkan dalam beberapa halaman kategori produk. Perlu adanya interaksi untuk mengganti dari satu halaman ke halaman berikutnya sehingga semua halaman yang menampilkan informasi produk dapat diambil datanya. Kode program untuk membuka halaman berikutnya pada halaman kategori produk Shopee adalah sebagai berikut.

```
while True:
    next_page = driver.find_elements_by_xpath(
        '//button[@class="shopee-button-outline shopee-mini-page-controller_next-btn"]')

    if (next_page):
        (get data)

        next_page[0].click()
    else:
        break
```

Pada halaman kategori produk terdapat dua informasi produk yang diambil, yaitu nama dan *URL* produk. Dalam menemukan lokasi elemen terhadap dua informasi tersebut dalam struktur halaman digunakan XPath. Selain informasi nama, dan *URL*, perlu diketahui dari kategori mana produk diambil sehingga dimasukkan nama kategori yang didapat dari proses pengambilan data sub-kategori. Konversi rancangan elemen data menjadi XPath pada halaman kategori produk Shopee dapat dilihat pada Tabel 5.5. Implementasi kode untuk *scraping* data produk dapat dilihat pada bagian LAMPIRAN Tabel A.6.

Tabel 5.5 Implementasi XPath Lokasi Elemen Data Produk Shopee.

Nama Atribut	XPath	Tipe Elemen
kategori	Didapatkan dari data URL kategori	
title	//*[@class="O6wiAW"]	<i>text</i>
url_produk	//*[@class="col-xs-2-4 shopee-search-item-result__item"]/div/a	<i>href</i>

C. *Scraping URL* UMKM Makanan dan Minuman Shopee

Data *URL* produk digunakan untuk mengakses halaman produk Shopee. Setiap *URL* produk diakses oleh *scraper* untuk mendapatkan *URL* UMKM. Akses *URL* UMKM digunakan *for* pada implementasi kode, dapat dilihat sebagai berikut.

```
for url in list_url_produk:
    (get url toko)
```

Pengambilan data *URL* UMKM dilakukan dengan melakukan konversi dari rancangan menjadi XPath. Implementasi XPath pada halaman produk Shopee untuk mengambil data *URL* UMKM dapat dilihat pada **Tabel 5.5**. Kode program untuk *scraping URL* UMKM di Shopee dapat dilihat pada bagian LAMPIRAN Tabel A.7.

Tabel 5.6 Rancangan Lokasi Elemen Data Produk Shopee.

Nama Atribut	XPath	Tipe Elemen
url_UMKM	//*[@class="_1jOO4S"]/a	<i>href</i>

D. *Scraping* Informasi UMKM Makanan dan Minuman Shopee

Setiap halaman UMKM diakses oleh *scraper* dengan menggunakan masukan data *URL* UMKM. Akses *URL* UMKM melalui kumpulan data dapat dilakukan menggunakan perulangan *for* pada Python. Pengulangan dalam kode untuk akses *URL* UMKM Shopee dapat dilihat sebagai berikut.

```
for url in list_url_toko:
    driver.get(url)

    (get data toko)
```

Pada halaman UMKM terdapat beberapa informasi yang dibutuhkan. Pengambilan data oleh *scraper* menggunakan XPath hasil konversi dari rancangan. Implementasi XPath untuk pengambilan data UMKM dapat dilihat pada Tabel 5.7. Setiap data yang diambil pada halaman UMKM disimpan dalam sebuah Dataframe seperti berikut. Implementasi *scraping* informasi UMKM Shopee dapat dilihat pada bagian LAMPIRAN Tabel A.8.

```
toko.loc[index toko] =
    ['name', 'link', 'city', 'seller_type',
     'product_category', 'numOf_product',
     'chat_performance', 'followers',
     'rating_review', 'since', 'voucher',
     'numOf_media', 'description']
```

Tabel 5.7 Implementasi XPath Lokasi Elemen Data UMKM Shopee.

Nama Atribut	XPath	Tipe Elemen
link	Didapatkan dari data <i>URL</i> UMKM	

name	//*[@class="section-seller-overview-horizontal__portrait-name"]	<i>text</i>
city	//*[@class="_3amru2"]	<i>text</i>
seller_type	//*[@class="ofs-header__page-name"]	<i>text</i>
product_category	//*[@class="shopee-category-list__body"]	<i>text</i>
numOf_product	//div[@class="section-seller-overview__item-text" and contains(., "produk")]/div[2]	<i>text</i>
chat_performance	//div[@class="section-seller-overview__item-text" and contains(., "Performa chat")]/div[2]	<i>text</i>
followers	//div[@class="section-seller-overview__item-text" and contains(., "pengikut")]/div[2]	<i>text</i>
rating_review	//div[@class="section-seller-overview__item-text" and contains(., "penilaian")]/div[2]	<i>text</i>
since	//div[@class="section-seller-overview__item-text" and contains(., "penilaian")]/div[2]	<i>text</i>
voucher	//*[@class="shop-page__section shop-page__vouchers"]	<i>boolean</i>
numOf_media	//*[@class="image-carousel__dots"]	<i>count</i>
description	//div[contains(@class, "shop-page-shop-description")]	<i>text</i>

5.1.3 Google My Business

Scraping di platform Google My Business membutuhkan beberapa data tambahan. Data tambahan yang dimaksud adalah data kecamatan/kelurahan, dan data kategori. Setelah data tambahan terpenuhi, maka dapat dilakukan *scraping* pada Google My Business.

Pada tahap ini dilakukan *scraping* data untuk mendapatkan informasi UMKM Google My Business yang relevan dengan kategori makanan dan minuman di Surabaya. Data pendukung, yaitu data kategori, serta kecamatan dan kelurahan digunakan sebagai masukan dalam penyusunan kata kunci pencarian. Kata kunci pencarian merupakan gabungan dari kategori dan kecamatan, misal *candy store* Kecamatan Sukolilo. Jumlah kata kunci pencarian yang digunakan untuk kategori makanan dan minuman adalah **73.150**. Implementasi penyusunan kata kunci pencarian pada kode dapat dilihat sebagai berikut.

```
list_keyword = []
for kategori in list_kategori_gb:
    for kecamatan in list_kecamatan:
        list_keyword.append(str(kategori) + ' ' +
                             str(kecamatan))
```

Scraping diawali dengan membuka halaman *google.com*, dan memasukkan kata kunci pencarian yang terdiri dari kategori dan kecamatan. *Driver* melakukan pencarian terhadap kata kunci yang dimasukkan. Implementasi pencarian dapat dilihat pada kode berikut.

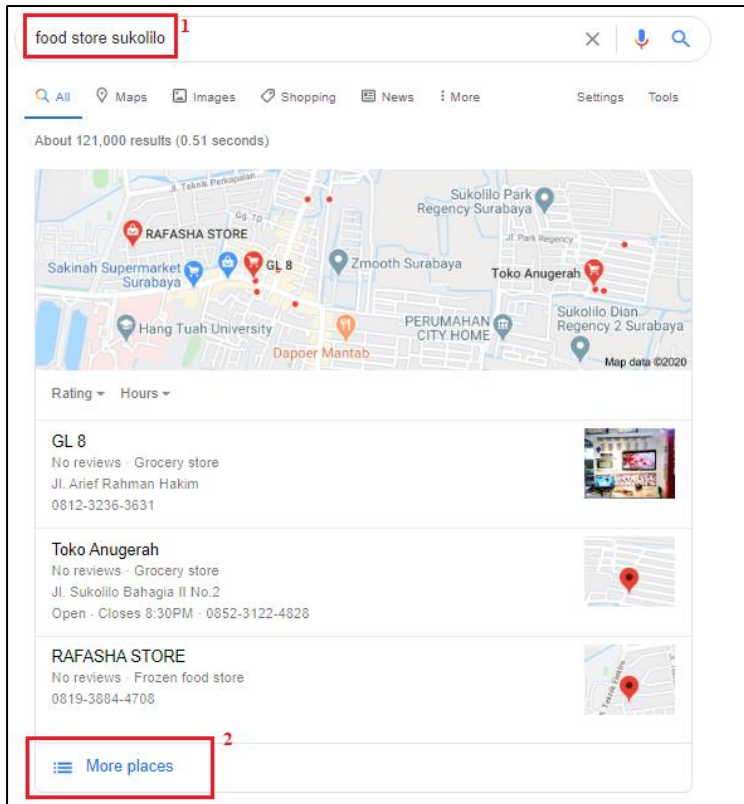
```
for keyword in list_keyword:
    search_element = driver.find_element_by_xpath(
        '//*[@class="gLFyf gsfi"]')

    search_element.clear()
    search_element.send_keys(keyword)
    search_element.send_keys(Keys.RETURN)
```

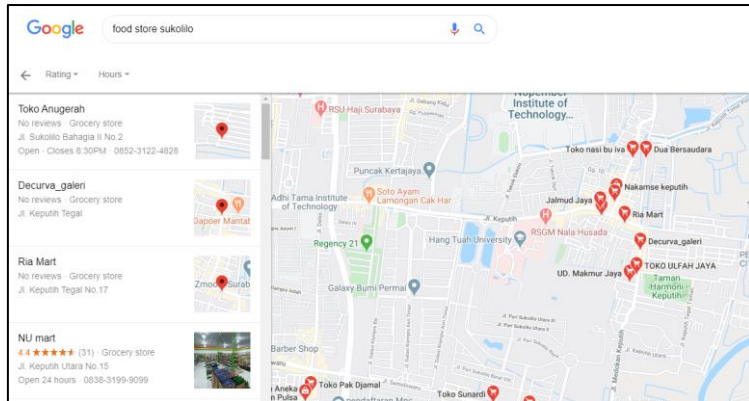
Pengulangan tiap kata kunci yang digunakan menggunakan *for* pada Python. XPath lokasi elemen *search box* adalah

`//*[@class="gLfyf_gsf"]`). Perintah `send_keys` untuk memasukkan kata kunci dalam *search box* seperti pada Gambar 5.1 dengan keterangan nomor 1. Memulai pencarian dengan kata kunci yang telah dimasukkan menggunakan perintah `send_keys` dengan parameter `Keys.RETURN` yang artinya sama dengan menekan tombol *Enter*. Masuk ke dalam halaman Google My Business apabila ditemukan hasil pencarian data kunci yang dicari dengan menekan tombol *more places* seperti pada Gambar 5.1 dengan keterangan nomor 2. Implementasi klik tombol *more places* dapat dilihat sebagai berikut.

```
morePlaces_element = driver.find_elements_by_xpath(  
    '//*[@class="zkIadb"]')  
  
morePlaces_element[0].click()
```

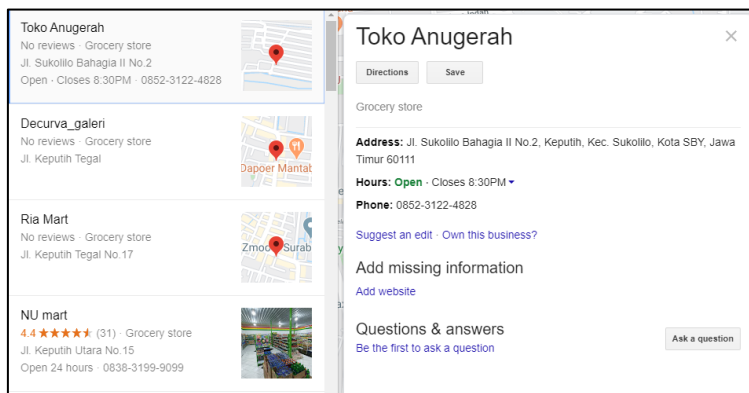


Gambar 5.1 Ilustrasi Hasil Pencarian Awal di google.com.



Gambar 5.2 Ilustrasi Halaman Pencarian di Google My Business.

Pada halaman hasil pencarian di Google My Business, *scraper* menemukan lokasi *container* yang berisi UMKM hasil pencarian. *Scraper* mengklik tiap baris UMKM pada hasil pencarian, sehingga menampilkan detail informasi UMKM seperti pada Gambar 5.3.



Gambar 5.3 Ilustrasi Detail Informasi UMKM Google My Business.

Google My Business memberikan beberapa informasi UMKM dari hasil pencarian yang akan diambil dan disimpan menjadi satu data, yaitu data UMKM Google My Business. Hasil pencarian di Google My Business terdiri dari beberapa Akses

elemen data UMKM Google My Business dilakukan dengan menggunakan XPath. Konversi rancangan elemen data UMKM Google My Business menjadi XPath dapat dilihat pada Tabel 5.8. Kode program untuk melakukan *scraping* data di Google My Business dapat dilihat pada Tabel A.9.

Tabel 5.8 Implementasi XPath Elemen Lokasi Data UMKM Google My Business.

Nama Atribut	XPath	Tipe Elemen
name	<code>//*[@class="SPZz6b"]/div/span</code>	<i>text</i>
website	<code>//*[@class="QqG1Sd"]/a</code>	<i>href</i>
rating	<code>//*[@class="Ob2kfd"]/div/span[1]</code>	<i>text</i>
numOf_google_review	<code>//*[@class="hqzQac"]/span/a/span</code>	<i>text</i>
price	<code>//*[@class="YhemCb"][0]</code>	<i>text</i>
category	<code>//*[@class="YhemCb"][1]</code>	<i>text</i>
address	<code>//*[@class="LrzXr"]</code>	<i>text</i>
operational_hours	<code>//div[@class="zloOqfPZPZlf"]/div/div/div[1]/div[2]/table/tbody</code>	<i>text</i>
phone	<code>//*[@class="LrzXr zdqRlf knofv"]</code>	<i>text</i>
instagram	<code>//*[@class="PZPZlf kno-vrt-t"]</code>	<i>href</i>

5.1.4 Instagram

Pertama-tama *scraper* membuka URL <https://www.instagram.com/instagram/>. Halaman Instagram memiliki dua perilaku. Perilaku pertama yaitu Instagram meminta untuk *login*. Maka, *scraper* perlu untuk memasukkan

username dan *password* menggunakan akun yang telah dibuat khusus untuk *scraping*. Dalam melakukan *login* dibuatlah fungsi seperti berikut.

```
try:
    username_element = driver.find_element_by_name('username')
    try:
        username_element.send_keys("scraper023")
    except:
        print('username field not found')
    try:
        password_element = driver.find_element_by_name('password')
        password_element.send_keys("scraper")
    except:
        print('password field not found')
    try:
        login_button_element = driver.find_element_by_xpath('//*[@class="sqdOP L3NKy y3zKF "]')
        login_button_element.click()
        print('logging in')
    except:
        print('login button not found')
except:
    print('no need to login')
```

Apabila halaman Instagram tidak meminta untuk *login*, maka *scraper* dapat memasukkan kata kunci pencarian ke dalam kolom pencarian Instagram. Dimulai dari menemukan elemen *search* hingga memasukkan nama UMKM ke dalam *field* pencarian. Implementasi pencarian di Instagram menggunakan kode dapat dilihat sebagai berikut.

```

search_element = driver.find_element_by_xpath(
    '//*[@class="LWmhU _0aCwM"]/input')
def searching():
    search_element.clear()
    search_element.send_keys(str(store))
    print('store ' + str(index) + ': ' + str(store))

```

5.2 Preprocess Data

5.2.1 Data UMKM Tokopedia

Pada *data preprocessing* Tokopedia proses *data cleaning* dan *data transformation* dilakukan pada seluruh atribut informasi UMKM tokopedia. Pada *data cleaning*, beberapa perlakuan seperti mengganti nilai kosong atau NAN dengan format '*no (nama atribut)*', dan menghapus duplikasi pada atribut data. Pada *data extraction*, data dilakukan ekstraksi menggunakan teks dari atribut *description*. Atribut baru yang dihasilkan dari proses ekstraksi yaitu *COD* (*Cash on Delivery*), *phone*, *website* dan *instagram*. *COD* memiliki arti *boolean* menunjukkan apakah UMKM memiliki dukungan layanan untuk bayar di tempat atau tidak.

Pada *attribute normalization*, beberapa data yang diubah adalah *followers* dengan merubah tipe data *string* menjadi *int*, merubah isi attribute *phone* agar sesuai dengan standard. Standar struktur *phone* yang dimaksud adalah "0 (nomor_operator) (nomor_telepon)". Lalu *attribute conversion*, mengonversikan sebuah atribut menjadi atribut lain seperti mengubah atribut *address* menjadi atribut baru yaitu kecamatan dan kelurahan. Secara keseluruhan detail atribut yang dilakukan *preprocess* dapat dilihat pada Tabel 5.9. Implementasi *preprocess* pada kode dapat dilihat pada LAMPIRAN Tabel A.10.

Tabel 5.9 *Preprocess* Data UMKM Tokopedia.

Nama Atribut	<i>Preprocess</i>
name	Tidak ada <i>preprocess</i>
link	Menghapus duplikasi data
seller_type	Mengganti <i>nan</i> dengan “Regular seller”
city	<i>Filter</i> area Surabaya
followers	Mengonversikan nilai ‘k’ dan ‘m’ menjadi <i>int</i>
reputation_point	Normalisasi dan membersihkan data
product_sold	<ul style="list-style-type: none"> - Mengkonversi nilai ‘k’ dan ‘m’ menjadi <i>int</i> - Mengganti nilai kosong menjadi nol (0)
rating	Tidak ada <i>preprocess</i>
review	Normalisasi dan membersihkan data
owner	Mengganti <i>value nan</i> dengan “no owner information”
description	Ekstraksi informasi menjadi atribut baru, yaitu COD, phone, website dan instagram
address	<ul style="list-style-type: none"> - Normalisasi dan membersihkan data - Mengonversikan atribut menjadi kecamatan dan kelurahan

since	Memisahkan atau <i>split</i> data menjadi atribut <i>monthly_joined</i> & <i>year_joined</i>
numOf_delivery_services	Tidak ada <i>preprocess</i>
delivery_services	Tidak ada <i>preprocess</i>
numOf_satisfied	Mengubah tipe data menjadi <i>int</i>
numOf_neutral	Mengubah tipe data menjadi <i>int</i>
numOf_unsatisfied	Mengubah tipe data menjadi <i>int</i>
process_speed	Tidak ada <i>preprocess</i>
numOf_storefront	Tidak ada <i>preprocess</i>
storefront	Tidak ada <i>preprocess</i>
active_product	<ul style="list-style-type: none"> - Menghapus baris data yang kosong - Mengubah tipe data menjadi <i>int</i>
success_transaction	Mengubah tipe data menjadi <i>int</i>
COD (hasil ekstrasi)	Tidak ada <i>preprocess</i>
phone (hasil ekstrasi)	Mengonversikan <i>value</i> atribut <i>phone</i> agar sesuai dengan standard
website (hasil ekstrasi)	Tidak ada <i>preprocess</i>
instagram (hasil ekstrasi)	Tidak ada <i>preprocess</i>
kecamatan (hasil ekstrasi)	Tidak ada <i>preprocess</i>
kelurahan (hasil ekstrasi)	Tidak ada <i>preprocess</i>

5.2.2 Data UMKM Shopee

Dilakukan ekstraksi informasi pada atribut *description* menjadi atribut baru yaitu *COD*, *phone*, *website* dan *instagram*. Dilakukan juga *attribute normalization* dan *attribute conversion* pada beberapa atribut. Secara keseluruhan detail atribut yang dilakukan *preprocess* dan kode program masing-masing dapat dilihat pada Tabel 5.10 dan LAMPIRAN Tabel A.11.

Tabel 5.10 *Preprocessing* Data Shopee.

Nama Atribut	<i>Preprocess</i>
Link	- Menjadikan sebagai index
Name	Menghapus <i>string</i> yang tidak memiliki standard <i>ascii</i>
City	Filter area Surabaya dengan menggunakan data kecamatan dan kelurahan, serta <i>string</i> “sby” dan “”
seller_type	Mengganti <i>nan</i> dengan “Regular seller”
product_category	Menghilangkan string ‘Semua Produk\n’
numOf_product	<ul style="list-style-type: none"> - Mengonversi nilai ‘RB’ menjadi <i>int</i> - Merubah tipe data menjadi <i>int</i>
chat_performance	<ul style="list-style-type: none"> - Memisahkan atau <i>split</i> data menjadi atribut <i>chat_performance_unit</i> - Konversi data persentase menjadi <i>int</i>

Followers	- Mengkonversi nilai 'RB' menjadi <i>int</i>
rating_review	- Memisah/ <i>split</i> data menjadi atribut <i>rating</i> dan <i>review</i> - Merubah tipe data menjadi <i>int</i> dan <i>float</i>
Since	Tidak ada <i>preprocess</i>
voucher	Tidak ada <i>preprocess</i>
numOf_media	Tidak ada <i>preprocess</i>
Description	Ekstraksi informasi menjadi atribut baru, yaitu <i>COD</i> , <i>phone</i> , <i>website</i> dan <i>instagram</i>
COD (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Phone (hasil ekstraksi)	Mengonversi isi atribut <i>phone</i> agar sesuai dengan standard nomor
Website (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Instagram (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Chat_performance_unit (hasil ekstraksi)	Tidak ada <i>preprocess</i>

5.2.3 Data UMKM Google My Business

Pada informasi UMKM google business juga dilakukan *data preprocessing* sesuai data yang didapatkan. Kegiatan *data cleaning* dengan mengganti dengan menghapus duplikasi data, mengganti data *NaN* atau data yang kosong dengan format "No (nama-atribut)". Pada *data transformation* beberapa perlakuan seperti, menambah atribut *operational_hours_duration* dari konversi data dari *operational_hours*.

Ekstraksi informasi dari atribut *website* yang mengandung atribut *e-commerce* menjadi atribut tokopedia, shopee dan instagram. Melakukan ekstraksi dan konversi informasi dari atribut *address* menjadi atribut kecamatan dan kelurahan. Secara keseluruhan atribut yang dilakukan *preprocess* dapat dilihat pada Tabel 5.11. Implementasi *script preprocess* Google My Business data dilihat pada LAMPIRAN Tabel A.12.

Tabel 5.11 *Preprocess* Data UMKM Google My Business.

Nama Atribut	<i>Preprocess</i>
Name	Menghapus data yang duplikat
Website	Ekstraksi informasi yang menjadi atribut baru : tokopedia. Shopee dan instagram
Rating	Tidak ada <i>preprocess</i>
numOf_google_review	Normalisasi dan membersihkan data
Price	Mengganti data yang kosong menjadi 'no price'
Category	Menghapus kategori yang tidak sesuai
Address	<ul style="list-style-type: none"> - Normalisasi dan membersihkan data - Konversi atribut menjadi kecamatan dan kelurahan
Operational_hours	Konversi atribut menjadi <i>operational_hours_duration</i>
Phone	Normalisasi dan membersihkan data

Instagram	Menambahkan data hasil ekstraksi informasi
Tokopedia	Normalisasi data
Shopee	Normalisasi data
Operational_hours_duration	Tidak ada <i>preprocess</i>
Kecamatan	Tidak ada <i>preprocess</i>
kelurahan	Tidak ada <i>preprocess</i>

5.3 Integrasi Data

Integrasi data dilakukan antara data UMKM Tokopedia dengan Shopee, dan Google My Business dengan Tokopedia dan Shopee. Integrasi dilakukan menggunakan metode *entity resolution* dengan teknik penghitungan *similarity* atau kemiripan. Data UMKM pada Google My Business, Tokopedia, Shopee, dan hasil *scraping* Instagram menjadi masukan untuk proses integrasi. Pertama-tama data UMKM dengan hasil *scraping* Instagram digabungkan menjadi satu sesuai dengan nama UMKM pada Tokopedia atau Shopee. Implementasi penggabungan data UMKM Tokopedia dan Shopee dengan hasil *scraping* Instagram dapat dilihat sebagai berikut.

```
df_tokped = pd.merge(df_tokped, df_ig_tokped, on='name', how='left')
df_shopee = pd.merge(df_shopee, df_ig_shopee, on='name', how='left')
```

Atribut data yang dihitung kemiripannya antara lain *phone*, *website*, *instagram* dan *name* karena masing-masing platform memiliki kesamaan pada keempat atribut tersebut. Data pada platform *marketplace* (Shopee dan Tokopedia) akan dicari kemiripan pada platform Google My Business. Kemiripan antar data menggunakan algoritma *sequencematcher* dengan perintah *ratio*. Dibuat sebuah fungsi bernama *similar* sebagai perintah khusus dalam melakukan *similarity* dengan parameter *a* atau teks pertama, dan *b* sebagai teks kedua. Implementasi algoritma *sequencematcher* pada kode dapat dilihat sebagai berikut.

```
def similar(a, b):
    return SequenceMatcher(None, a.lower(), b.lower()).ratio()
```

Pengukuran nilai *similarity* dilakukan dengan melakukan perulangan untuk setiap UMKM Tokopedia dan Shopee terhadap setiap took di Google My Business. Dilakukan pengukuran kemiripan terhadap empat atribut yang sama antar platform, yaitu *phone*, *website*, *instagram*, dan *name*.

Setiap pengecekan atribut untuk *similarity* diambil *similarity* dengan hasil paling maksimal untuk setiap UMKM Tokopedia dan Shopee terhadap Google My Business dengan disimpan sementara pada atribut *ratio_max*. Digunakan sebuah *threshold* atau nilai minimum hasil *similarity* sehingga sebuah UMKM di Tokopedia dan Shopee dapat dikatakan sama dengan UMKM di Google My Business. *Similarity* memiliki urutan prioritas pengecekan atribut dimulai dari *phone*, kemudian apabila hasil *similarity* bernilai minimum *threshold* yang telah ditentukan maka dikatakan sebuah UMKM yang sama. Apabila memiliki nilai kurang dari *threshold* maka lanjut dilakukan pengecekan atribut prioritas berikutnya, yaitu *website*, *instagram*, dan *name*. Penentuan nilai *threshold* dilakukan dengan melakukan eksperimen pada setiap atribut yang digunakan untuk penghitungan *similarity*. Dengan melakukan validasi manual, setiap kali menentukan *threshold* dihitung berapa banyak yang memiliki nilai *True Positive* (pasti benar) dengan nilai *similarity* 1 dibandingkan dengan jumlah UMKM yang dianggap sama (memiliki nilai *ratio* lebih dari sama dengan *threshold*). Didapatkan nilai *threshold* untuk masing-masing atribut yang digunakan untuk minimal nilai *similarity* dapat dilihat pada Tabel 5.12. Implementasi penghitungan hasil *similarity* dapat dilihat sebagai berikut.

```

similarity = similar(phone_test,list_phone_from_gb[idx_gb])
if similarity > ratio_max:
    idx_ratio_max = idx_gb
    ratio_max = similarity

if ratio_max >= threshold:
    list_tokopedia_gb[idx_ratio_max] = list_link_from_tokped[idx_tokped]
    list_tokopedia_similarity[idx_ratio_max] = ratio_max
    list_tokopedia_similarity_by[idx_ratio_max] = 'atribut'

```

Secara keseluruhan, terdapat dua integrasi. Integrasi pertama antara Tokopedia dengan Shopee dapat dilihat pada bagian LAMPIRAN Tabel A.13. Integrasi kedua dilakukan antara Google My Business dengan Tokopedia dan Shopee, kode program dapat dilihat pada LAMPIRAN Tabel A.14.

Tabel 5.12 *Threshold* Nilai *Similarity* Tiap Atribut.

Prioritas	Nama Atribut	<i>Threshold</i>
1	phone	0.96
2	website	0.86
3	instagram	0.83
4	name	0.80

5.4 Perhitungan Nilai Tingkat Kebenaran

Nilai presisi dihitung untuk mendapatkan tingkat kualitas data. Perhitungan nilai presisi dilakukan dengan melakukan pengujian dengan beberapa tahap, yaitu menentukan nilai *confidence level*, menghitung nilai sampel, mengambil sampel, dan melakukan pengujian. Dari hasil pengujian yang dapat dilihat pada Tabel 5.13, maka digunakan *threshold* dengan tingkat presisi paling tinggi, yaitu **0.8**.

Tabel 5.13 Hasil Pengujian.

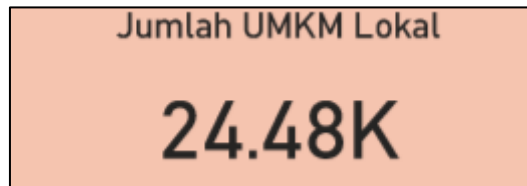
<i>Threshold</i>	Jumlah data yang berhasil digabungkan	Jumlah sampel	Tingkat kebenaran (Presisi)
0.6	2441	333	0.6324
0.8	1051	291	0.6956

5.5 Visualisasi dan Analisis

Implementasi visualisasi dilakukan berdasarkan matriks yang telah dibuat. Rincian visualisasi dibagi menjadi Card, Gauge, Scatter Plot, Bar Chart, donut chart, Wordcloud dan Map.

5.5.1 Card

Grafik visualisasi data berupa *card* digunakan untuk menunjukkan suatu data bersifat numerik secara ringkas dan jelas. Dalam penelitian ini, salah satu implementasi *card* digunakan untuk menampilkan informasi *single*, seperti jumlah UMKM. Contoh hasil implementasi visualisasi Card dapat dilihat pada Gambar 5.4.

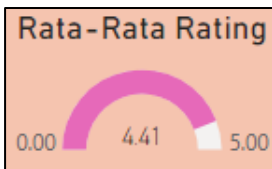


Gambar 5.4 Implementasi Visualisasi Card.

5.5.2 Gauge

Gauge dalam visualisasi data merupakan jenis grafik yang mematerialisasi dimana skala yang digunakan mewakili matriks, *pointer* mewakili dimensi, dan sudut *pointer* merepresentasikan nilai. Grafik ini biasa digunakan untuk menampilkan ketercapaian dari sebuah indikator dan

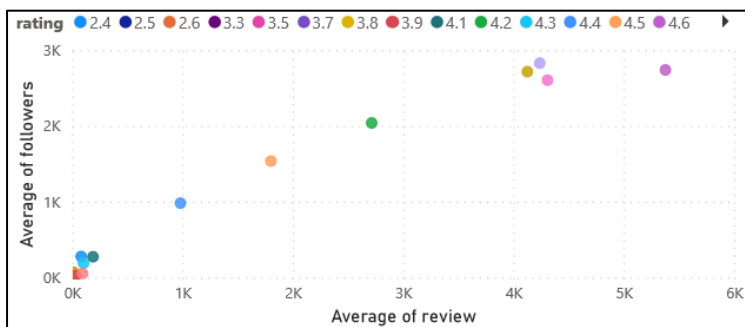
membandingkan interval. Dalam penelitian ini, salah satu implementasi *card* digunakan untuk menampilkan informasi nilai skala, seperti *rating* dengan skala 1 sampai 5. Contoh implementasi Gauge dapat dilihat pada Gambar 5.5.



Gambar 5.5 Implementasi Visualisasi *Gauge*.

5.5.3 Scatter Plot

Scatter plot atau dapat disebut dengan diagram pencar adalah grafik visualisasi yang digunakan untuk menunjukkan hubungan (korelasi) antar dua variabel. Contoh *implementasi scatter plot* dalam penelitian ini adalah untuk mengetahui korelasi antara rata-rata jumlah *followers* dan rata-rata jumlah *review*. Implementasi penggunaan scatter plot dapat dilihat pada Gambar 5.6.

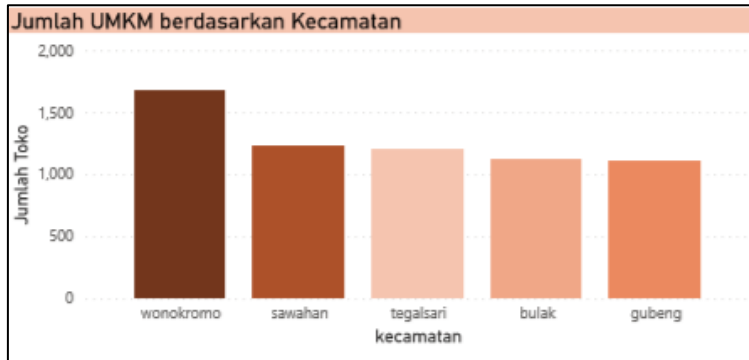


Gambar 5.6 Implementasi Visualisasi *Scatter Plot*.

5.5.4 Bar Chart

Bar chart atau diagram batang merupakan grafik berbentuk batang yang biasa digunakan untuk memvisualisasikan perbandingan data. Salah satu contoh visualisasi perbandingan data yang dilakukan dalam penelitian

ini adalah membandingkan jumlah UMKM berdasarkan kecamatan. Contoh implementasi visualisasi *bar chart* dapat dilihat pada Gambar 5.7.



Gambar 5.7 Implementasi Visualisasi *Bar Chart*.

5.5.5 Map

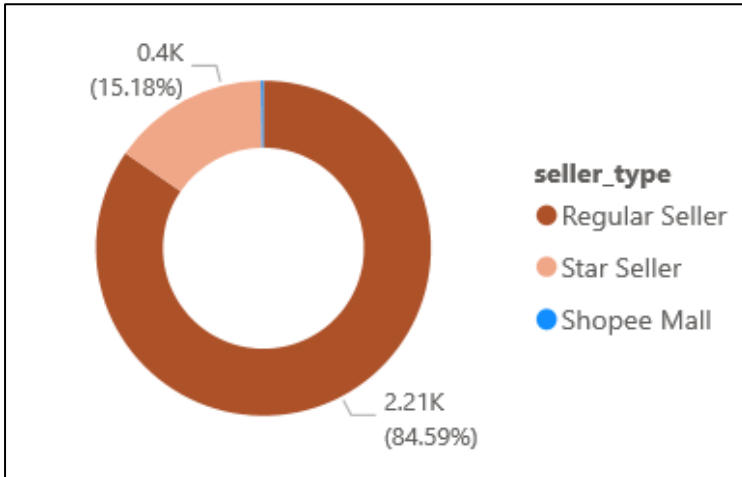
Visualisasi *map* digunakan untuk menunjukkan data dengan unsur geografis sehingga data dapat diinterpretasikan sesuai dengan lokasi. Salah satu contoh *map* digunakan untuk menampilkan informasi persebaran jumlah UMKM seperti pada Gambar 5.8.



Gambar 5.8 Implementasi Visualisasi *Map*.

5.5.6 Donut Chart

Donut chart merupakan variasi dari *pie chart* yang memiliki lubang pada tengah grafik. Visualisasi ini berguna untuk menunjukkan persentase pembagian dari keseluruhan atau proporsi nilai dalam data. Proporsi data dapat ditunjukkan seperti pada Gambar 5.9 dengan menunjukkan proporsi jumlah UMKM berdasarkan jenis *seller*.



Gambar 5.9 Implementasi Visualisasi *Donut Chart*.

5.5.7 Wordcloud

Wordcloud digunakan dalam memvisualisasikan kata-kata dengan menyoroti frekuensi kata yang digunakan dalam teks tertulis. *Wordcloud* dapat digunakan salah contohnya untuk menampilkan informasi kata yang sering disebut pada data kategori makanan dan minuman. Visualisasi wordcloud dapat dilihat pada Gambar 5.10.

Gambar 5.10 Implementasi Visualisasi Wordcloud.

Halaman ini sengaja dikosongkan.

BAB VI

HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil dari setiap proses dari implementasi. Mulai dari hasil *scraping* data, visualisasi, hingga analisis.

6.1 Data

Hasil akhir dari *scraping data* adalah data yang berisi informasi UMKM pada tiap platform, yaitu Tokopedia, Shopee, dan Google Business. Mulai dari tahap mendapatkan informasi kategori hingga informasi UMKM.

6.1.1 Tokopedia

Hasil dari *scraping* pada platform Tokopedia adalah data kategori, *URL* dan nama produk, *URL* UMKM, dan informasi UMKM.

A. Data Informasi Kategori Makanan dan Minuman Tokopedia

Kategori makanan dan minuman di Tokopedia berjumlah sebanyak **147**. Data diambil dari sub-kategori dan sub-sub kategori pada makanan dan minuman. Pratinjau data kategori makanan dan minuman pada Tokopedia dapat dilihat pada Gambar 6.1. Data kategori makanan dan minuman secara keseluruhan dapat dilihat pada LAMPIRAN Tabel B.1.

	kategori	link
0	bahan kue	https://www.tokopedia.com/p/makanan-minuman/ba...
1	beras	https://www.tokopedia.com/p/makanan-minuman/beras
2	bumbu & bahan masakan	https://www.tokopedia.com/p/makanan-minuman/bu...
3	kopi	https://www.tokopedia.com/p/makanan-minuman/kopi
4	kue	https://www.tokopedia.com/p/makanan-minuman/kue
...
142	tepung kentang	https://www.tokopedia.com/p/makanan-minuman/te...
143	tepung roti - panir	https://www.tokopedia.com/p/makanan-minuman/te...
144	tepung sagu	https://www.tokopedia.com/p/makanan-minuman/te...
145	tepung tapioka	https://www.tokopedia.com/p/makanan-minuman/te...
146	tepung terigu & gandum	https://www.tokopedia.com/p/makanan-minuman/te...
147 rows x 2 columns		

Gambar 6.1 Pratinjau Data Kategori Makanan dan Minuman Tokopedia.

B. Data Informasi Produk Makanan dan Minuman Tokopedia

Scraping informasi produk makanan dan minuman Tokopedia menghasilkan data produk dengan atribut kategori, *title* atau nama produk, dan *link* atau *URL* produk. Pratinjau data produk makanan dan minuman di Tokopedia dapat dilihat pada Gambar 6.2.

	kategori	title	link
	bahan kue	VANILLA BEAN PREMIUM 10gr (Biji Vanilla) or VA...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
	bahan kue	Edible Picture/Edible Stamp/Stiker Makanan Min...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
	bahan kue	Edible Picture/Edible Stamp/Stiker Makanan Min...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
	bahan kue	Edible Stamp/Edible Picture/Stiker Makanan Min...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
	bahan kue	Edible Stamp/Edible Picture/Stiker Makanan Min...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...

	tepung terigu & gandum	Disk mill mesin penepung serbaguna 2 in 1 pene...	https://www.tokopedia.com/agromesin/disk-mill...
	tepung terigu & gandum	Rostitas Burrito Tortilla New Softer Wraps 460g...	https://www.tokopedia.com/purleshop/rostitas-bu...
	tepung terigu & gandum	TEPUNG BIANG KFC TEPUNG FRIED CHICKEN 1KG	https://www.tokopedia.com/purleshop/tepung-bia...
	tepung terigu & gandum	Mie Mentah Keriting Special Telur Bebek Untuk ...	https://www.tokopedia.com/purleshop/mie-mentah...
	tepung terigu & gandum	mr.rainboo spiku	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
254862 rows x 2 columns			

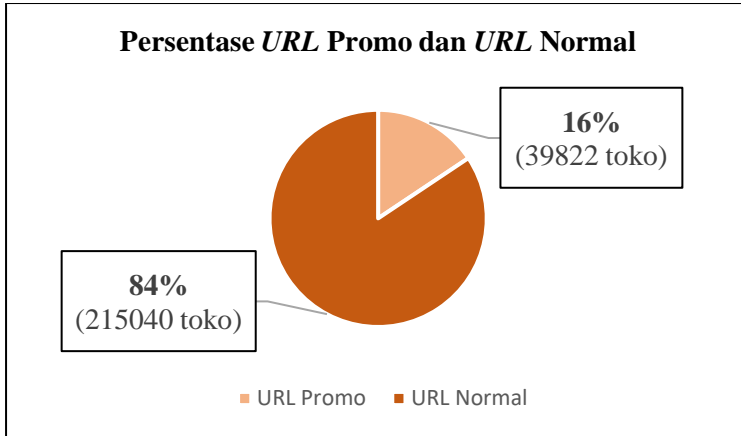
Gambar 6.2 Pratinjau Data Produk Kategori Makanan dan Minuman Tokopedia.

Secara keseluruhan data produk makanan dan minuman Tokopedia dapat dilihat pada LAMPIRAN Tabel B.2. Dilakukan analisis *wordcloud* pada nama produk makanan dan minuman Tokopedia sehingga didapatkan tren produk yang dijual berdasarkan nama produk seperti pada Gambar 6.3.



Gambar 6.3 *Wordcloud* Nama Produk Makanan dan Minuman Tokopedia.

Terlihat bahwa produk yang paling banyak dijual adalah susu kambing serta terkait dengan kopi, dan *green tea*. Data produk yang berhasil didapatkan adalah **254.862** produk beserta *URL* yang terdiri dari **39.822** *URL* produk promo, dan **215.040**. Persentase perbandingan jumlah *URL* promo dan normal dapat dilihat pada Gambar 6.4.



Gambar 6.4 Diagram Persentase Perbandingan *URL* Promo dan *URL* Normal.

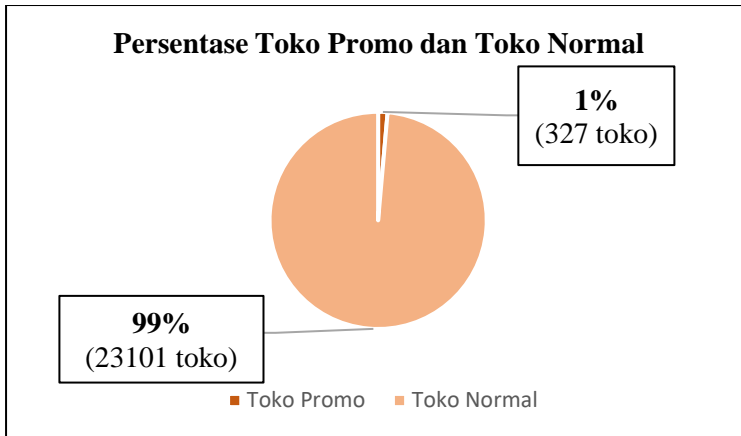
C. Data *URL* UMKM Makanan dan Minuman Tokopedia

Dari hasil *preprocess URL* produk menjadi *URL* UMKM serta penghilangan duplikasi didapatkan data sebanyak **23.428** *URL* UMKM di mana terdapat **327** UMKM didapatkan dari *URL* produk dengan promo, dan **23.101** tanpa promo. Pratinjau data *URL* UMKM makanan dan minuman Tokopedia dapat dilihat pada Gambar 6.5.

	link
0	https://www.tokopedia.com/healthynoon
1	https://www.tokopedia.com/jitron
2	https://www.tokopedia.com/realfood
3	https://www.tokopedia.com/dhanishop
4	https://www.tokopedia.com/waletkuindonesia
...	...
23423	https://www.tokopedia.com/eskrim
23424	https://www.tokopedia.com/teguhpatricia
23425	https://www.tokopedia.com/rereys
23426	https://www.tokopedia.com/globalmandirisby
23427	https://www.tokopedia.com/gonoos
23428 rows × 1 columns	

Gambar 6.5 Pratinjau Data *URL* UMKM Makanan Minuman Tokopedia.

Secara keseluruhan, data *URL* UMKM dapat dilihat pada LAMPIRAN Tabel B.3. Persentase perbandingan jumlah UMKM yang didapatkan dari produk promo dengan produk normal dapat dilihat pada Gambar 6.6.



Gambar 6.6 Diagram Persentase Perbandingan UMKM Promo dan UMKM Normal.

D. Data Informasi UMKM Makanan dan Minuman Tokopedia

Scraping informasi UMKM makanan dan minuman di Tokopedia menghasilkan data sebanyak **23.428** sesuai dengan jumlah *URL* UMKM yang didapatkan. Pratinjau data informasi UMKM hasil *scraping* dapat dilihat pada Gambar 6.7. Keseluruhan data UMKM makanan dan minuman Tokopedia dapat dilihat di LAMPIRAN Tabel B.9.

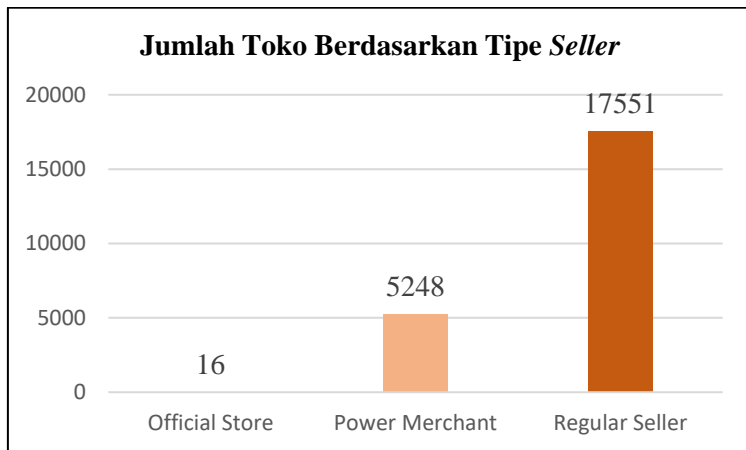
name	link	seller_type	city	followers	reputation_point	product_sold	rating	review	owner	...	run
0 Healthy Noona	https://www.tokopedia.com/healthynoonaa	Power Merchant	Kota Surabaya	1.7K Followers	16.123 points	23.7K	4.7	(4162 Ulasan)	Ivan Kivang
1 JITRON INDONESIA	https://www.tokopedia.com/jitron	Power Merchant	Kota Surabaya	2 Followers	NaN	3	5.0	(3 Ulasan)	Emanuel Vianney
2 Realfood	https://www.tokopedia.com/realfood	Official Store	Kota Surabaya	1.1K Followers	740 points	747	4.9	(131 Ulasan)	Realfood
3 Diani shop	https://www.tokopedia.com/dianishop	Power Merchant	Kota Surabaya	NaN	2 points	1	5.0	(1 Ulasan)	Ila Rafani
4 Gunung Walest Indonesia	https://www.tokopedia.com/walestindonesia	Power Merchant	Kota Surabaya	56 Followers	977 points	994	4.9	(204 Ulasan)	Hery
...
23423 Ice Cream Appliances	https://www.tokopedia.com/ieskrim	NaN	Surabaya	5 Followers	0 points	5	NaN	NaN	Alvia
23424 TP Jaya	https://www.tokopedia.com/teguhpatriana	NaN	Surabaya	1 Follower	0 points	NaN	NaN	NaN	Tegeh Playitno
23425 Rereys	https://www.tokopedia.com/rereys	NaN	Dukuh Pasis	NaN	4 points	80	NaN	NaN	kunijoro muniratul
23426 global mandiri surabaya	https://www.tokopedia.com/globalmandiristby	NaN	Surabaya	NaN	0 points	NaN	NaN	NaN	Yanuar
23427 Goneses	https://www.tokopedia.com/goneses	NaN	Dukuh Pasis	NaN	0 points	52	NaN	NaN	haripono tedhastur

Gambar 6.7 Pratinjau Hasil *Scraping* Data Informasi UMKM Makanan dan Minuman Tokopedia.

Data dari hasil *scraping* dilakukan *preprocess* sehingga data UMKM siap untuk digunakan analisis. *Preprocess* yang dilakukan adalah dengan menghapus data UMKM yang bukan berasal dari Surabaya, sehingga didapatkan data hasil *preprocess* sebanyak **22.815**.

Berdasarkan analisis dan *preprocess* yang dilakukan di tiap atribut data UMKM, berikut adalah fakta-fakta pada data UMKM makanan dan minuman Tokopedia.

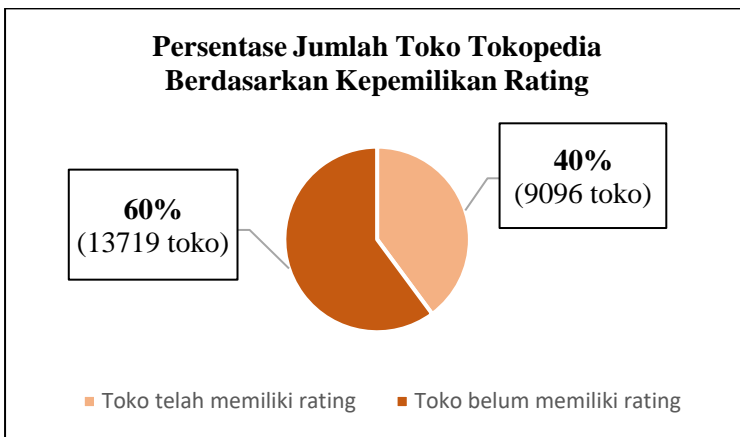
- a) Jumlah UMKM terbanyak berdasarkan tipe penjual adalah 17551 oleh *Regular Seller*, sedangkan terendah adalah 16 oleh *Official Store*. Persentase jumlah UMKM makanan dan minuman Tokopedia berdasarkan tipe penjual dapat dilihat pada Gambar 6.8.



Gambar 6.8 Jumlah UMKM Berdasarkan Tipe *Seller*.

- b) Rata-rata jumlah *review* UMKM makanan dan minuman di Tokopedia adalah **152** dengan **13.717** UMKM tidak memiliki *review* sama sekali.
- c) Rata-rata *reputation_point* yang dimiliki UMKM makanan dan minuman Tokopedia adalah **390,98** dengan **12.881** UMKM belum memiliki *reputation_point*, dan **21** UMKM memiliki *reputation_point* negatif.

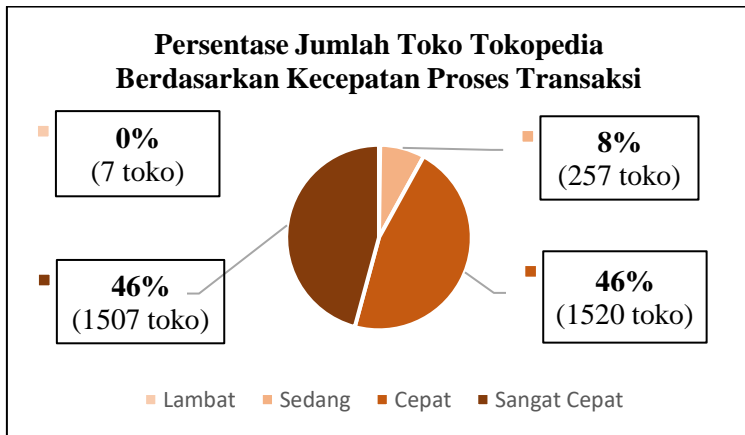
- d) Jumlah produk makanan minuman terjual di Tokopedia keseluruhan sebanyak **59.470.131** unit produk. Rata-rata produk terjual adalah **2.607** dengan **11.398** UMKM belum menjual produk sama sekali.
- e) Dari **22.815** UMKM terdapat **13.719** UMKM yang belum memiliki *rating*, dan **9.096** UMKM telah memiliki *rating*. Persentase perbandingan jumlah UMKM yang telah dengan belum memiliki *rating* dapat dilihat pada Gambar 6.9. Dari 9.096 UMKM yang telah memiliki *rating*, didapatkan rata-rata *rating* UMKM makanan dan minuman Tokopedia adalah **4.7**. Dari data tersebut masih banyak UMKM yang belum mendapatkan penilaian, tetapi untuk UMKM yang telah mendapatkan penilaian mendapatkan rata-rata penilaian yang cukup tinggi.



Gambar 6.9 Persentase UMKM yang Telah Memiliki *Rating* dengan yang Belum Memiliki *Rating*.

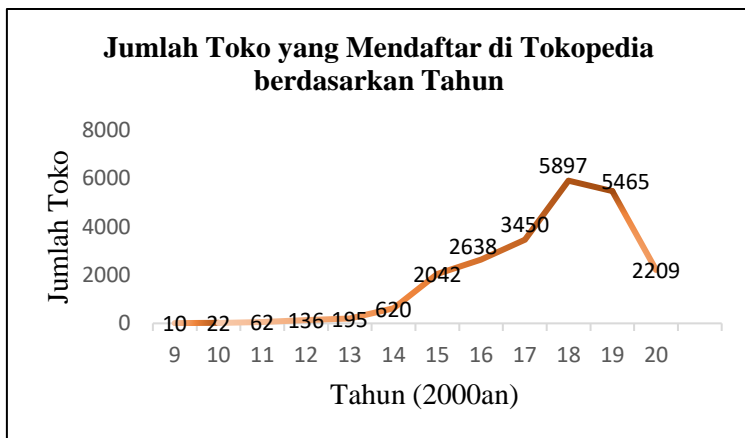
- f) Terdapat 3.291 UMKM yang telah memiliki kecepatan proses transaksi, sedangkan 19.524 UMKM masih belum informasi mengenai kecepatan proses transaksi. Dari **3.291** UMKM yang memiliki informasi kecepatan proses transaksi didapatkan paling banyak, yaitu sebanyak **1.520** UMKM termasuk cepat. Persentase jumlah toko di

Tokopedia berdasarkan kecepatan proses transaksi dapat dilihat pada Gambar 6.10.



Gambar 6.10 Persentase Jumlah UMKM Tokopedia Berdasarkan Kecepatan Proses Transaksi.

- g) Tahun paling banyak UMKM mendaftar adalah tahun 2018 dengan jumlah **5897** UMKM. Jumlah UMKM yang mendaftar pada platform Tokopedia berdasarkan tahun dapat dilihat pada Gambar 6.11.



Gambar 6.11 Jumlah UMKM Makanan Minuman yang Mendaftar di Tokopedia Berdasarkan Tahun.

6.1.2 Shopee

Scraping pada platform Shopee menghasilkan data kategori, produk, dan UMKM. Masing-masing data dijelaskan sebagai berikut.

A. Data Informasi Kategori Makanan dan Minuman Tokopedia

Data informasi kategori makanan dan minuman yang didapatkan dari Shopee berjumlah **12**. Pratinjau data hasil *scraping* kategori makanan dan minuman di Shopee dapat dilihat pada Gambar 6.12. Keseluruhan data kategori makanan dan minuman Shopee dapat dilihat pada LAMPIRAN Tabel B.5.

	kategori	link
0	makanan & minuman	https://shopee.co.id/Makanan-Minuman-cat.157
1	menu sarapan	https://shopee.co.id/Menu-Sarapan-cat.157.14953
2	roti & kue	https://shopee.co.id/Roti-Kue-cat.157.18449
3	susu & olahan	https://shopee.co.id/Susu-Olahan-cat.157.14574
4	bahan pokok	https://shopee.co.id/Bahan-Pokok-cat.157.14560
5	makanan instan	https://shopee.co.id/Makanan-Instan-cat.157.12561
6	makanan kaleng	https://shopee.co.id/Makanan-Kaleng-cat.157.18437
7	makanan siap saji	https://shopee.co.id/Makanan-Siap-Saji-cat.157...
8	makanan ringan	https://shopee.co.id/Makanan-Ringan-cat.157.14549
9	makanan beku	https://shopee.co.id/Makanan-Beku-cat.157.12567
10	cokelat & permen	https://shopee.co.id/Cokelat-Permen-cat.157.14582
11	minuman	https://shopee.co.id/Minuman-cat.157.1445
12	makanan segar	https://shopee.co.id/Makanan-Segar-cat.157.18681
13	makanan & minuman lainnya	https://shopee.co.id/Makanan-Minuman-Lainnya-c...

Gambar 6.12 Pratinjau Data Kategori Makanan dan Minuman Shopee.

B. Data Informasi Produk Makanan dan Minuman Shopee

Hasil *scraping* informasi produk makanan dan minuman didapatkan data sebanyak **19920** produk. Pratinjau data produk dapat dilihat pada Gambar 6.13.

	kategori	title	link
0	makanan & minuman	Sarang Madu Melifera 500g	https://shopee.co.id/Sarang-Madu-Melifera-500g...
1	makanan & minuman	Mie lidi siumang umang snack mi lidi pedas jaj...	https://shopee.co.id/Mie-lidi-siumang-umang-sn...
2	makanan & minuman	Bronchips Keripik Brownies Keju Kacang Greante...	https://shopee.co.id/Bronchips-Keripik-Brownie...
3	makanan & minuman	Makaroni Ngocor	https://shopee.co.id/Makaroni-Ngocor-i-7909176...
4	makanan & minuman	[GROSIR 1750]Ngocor Cemilan Instan MAKARONI NG...	https://shopee.co.id/-GROSIR-1750-Ngocor-Cemil...
...
19915	makanan & minuman lainnya	Good Time cookies isi 12pc DoublechocinGrosir	https://shopee.co.id/Good-Time-cookies-isi-12p...
19916	makanan & minuman lainnya	Mango Guava Pasion Fruit 60 mili WLCM Fr0z3n P...	https://shopee.co.id/Mango-Guava-Pasion-Fruit...
19917	makanan & minuman lainnya	Jual Starter Mocaf Bimo CF Tepung Cassava Ecer...	https://shopee.co.id/Jual-Starter-Mocaf-Bimo-C...
19918	makanan & minuman lainnya	Daltsabu mini chikuwa 1kg - F0339	https://shopee.co.id/Daltsabu-mini-chikuwa-1kg...
19919	makanan & minuman lainnya	SARANG MADU ASLI / HONEYCOMB COD	https://shopee.co.id/SARANG-MADU-ASLI-HONEYCOM...

19920 rows x 3 columns

Gambar 6.13 Pratinjau Data Produk Makanan dan Minuman Shopee.

Keseluruhan data produk makanan dan minuman Shopee dapat dilihat pada LAMPIRAN Tabel B.6. Didapati bahwa jumlah data produk aygn didapatkan dari kedua belas kategori makanan dan minuman di Shopee masing-masing sebanyak **1485**, kecuali pada kategori makanan kaleng sebanyak **615** produk. Hal tersebut terjadi karena Shopee membatasi jumlah produk yang ditampilkan pada masing-masing kategori. Jumlah data masing-masing kategori dapat dilihat pada Gambar 6.14.

	title
kategori	
bahan pokok	1485
cokelat & permen	1485
makanan & minuman	1485
makanan & minuman lainnya	1485
makanan beku	1485
makanan instan	1485
makanan kaleng	615
makanan ringan	1485
makanan segar	1485
makanan siap saji	1485
menu sarapan	1485
minuman	1485
roti & kue	1485
susu & olahan	1485

gagal dimuat, sehingga didapatkan data yang korup. Pratinjau data *URL* UMKM makanan dan minuman Shopee dapat dilihat pada Gambar 6.16. Keseluruhan data *URL* UMKM makanan dan minuman Shopee dapat dilihat pada LAMPIRAN Tabel B.7.

	link
0	https://shopee.co.id/ferhat.kevin.novianti.seller
1	https://shopee.co.id/emanwar
2	https://shopee.co.id/emanwar
3	https://shopee.co.id/rachelsrein44
4	https://shopee.co.id/emanwar
...	...
19840	https://shopee.co.id/insya1malik
19841	https://shopee.co.id/angelinesantoso
19842	https://shopee.co.id/kebunonline
19843	https://shopee.co.id/pt.masuyadistra
19844	https://shopee.co.id/24widdy_
19845 rows × 1 columns	

Gambar 6.16 Pratinjau Data *URL* UMKM Makanan dan Minuman Shopee.

D. Data Informasi UMKM Makanan dan Minuman Shopee

Scraping informasi UMKM makanan dan minuman Shopee menghasilkan data sebanyak **6.913** baris. Pratinjau data UMKM Shopee hasil *scraping* dapat dilihat pada Gambar 6.17.

	name	link	city	seller_type	product_category	numOf_product	chat_performance	follow
0	Omah Madu Malang	https://shopee.co.id/ferhat.kevin.novianti.seller	KAB. MALANG	NaN	Semua ProdukMadu & Clesan	2	51% (hitungan Menit)	
1	Mas Shop	https://shopee.co.id/emanwar	KOTA SURABAYA	Star Seller	Semua ProdukBaso acinJajananFroze Foodin...	374	96% (hitungan Jam)	77
2	RN Snacks	https://shopee.co.id/rachelsreln44	KOTA SURABAYA	Star Seller	Semua ProdukNew Arrivalspacking ananinCami...	170	93% (hitungan Jam)	16,8
3	Arimbi_Arimbi	https://shopee.co.id/arimbi_arimbi	KAB. SIDOARJO	Star Seller	Semua ProdukMakanan InstaninPerfengkapan Rum...	270	100% (hitungan Jam)	1
4	grestadini	https://shopee.co.id/grestadini	KAB. SIDOARJO	Star Seller	Semua ProdukMakanan Ringan LainnyainGemlan...	22	96% (hitungan Jam)	1
...
6908	gemmaroom	https://shopee.co.id/gemmaroom	KAB. SIDOARJO	NaN	Semua ProdukMakanan Siap SajiinMakanan Beku...	5.0	...% (hitungan Menit)	
6909	suolindah6	https://shopee.co.id/suolindah6	KOTA SURABAYA	NaN	Semua ProdukMakanan Beku Lainnya	4.0	94% (hitungan Hari)	4
6910	rindangaulaa	https://shopee.co.id/rindangaulaa	KAB. PACITAN	NaN	Semua ProdukMakanan Beku LainnyainPerawatan...	10.0	87% (hitungan Menit)	1
6911	damayshopolic	https://shopee.co.id/damayshopolic	KAB. SIDOARJO	NaN	Semua ProdukTas WanitaMakanan Beku inPerawa...	20.0	...% (hitungan Jam)	
6912	alfatana_idea	https://shopee.co.id/asrori88	KAB. SITUBONDO	NaN	ProdukRENGINANinKOP BUBUKinKERUPUK...	30.0	66% (hitungan Jam)	2

6913 rows x 13 columns

Gambar 6.17 Pratinjau Data UMKM Shopee Hasil *Scraping*.

Dari hasil *scraping*, masih didapatkan banyak UMKM yang bukan berada di Surabaya karena *filter* lokasi Shopee yang digunakan adalah Jawa Timur. Dilakukan *preprocess* lokasi kota UMKM dan penghapusan duplikasi *URL*, sehingga didapatkan UMKM makanan dan minuman yang berada di Surabaya pada Shopee sebanyak **2.609**. Pratinjau data UMKM makanan dan minuman Shopee dapat dilihat pada Gambar 6.18. Keseluruhan data UMKM makanan dan minuman Shopee dapat dilihat pada LAMPIRAN Tabel B.10.

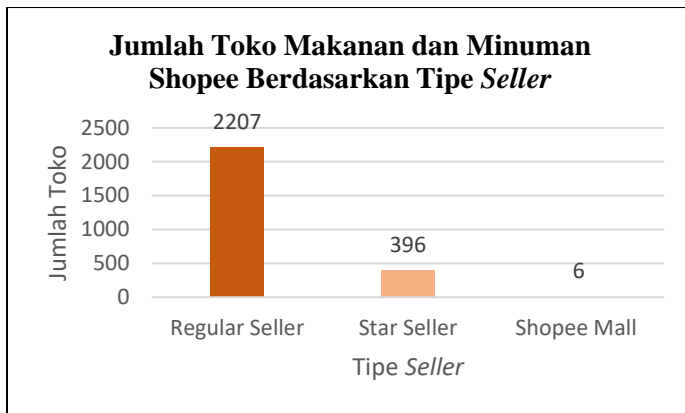
	name	link	seller_type	city	followers	rating	review	chat_performance_percentage	chat_performance_unit
0	Mas Shop	https://shopee.co.id/emanivar	Star Seller	kota surabaya	77000.0	4.8	86300.0	0.96	Jam
1	RN Snacks	https://shopee.co.id/rachelsrein44	Star Seller	kota surabaya	16800.0	4.8	24800.0	0.93	Jam
2	dibumbuin	https://shopee.co.id/dibumbuin	Star Seller	kota surabaya	9500.0	4.8	12900.0	0.98	Jam
3	Depot Bu Rudy Official Shop	https://shopee.co.id/depotburudyofficial	Shopee Mall	kota surabaya	24000.0	4.9	21900.0	0.98	Jam
4	LOKKEN STORE	https://shopee.co.id/sheashaquair16	Star Seller	kota surabaya	1700.0	4.7	2400.0	0.98	Menit
...
2604	RizieK_Shibab	https://shopee.co.id/rizieK_shibab	Regular Seller	kota surabaya	39.0	NaN	133.0	0.91	Menit
2605	nobelirfan	https://shopee.co.id/nobelirfan	Regular Seller	kota surabaya	29.0	NaN	1.0	NaN	NaN
2606	ms.pyurple	https://shopee.co.id/ms.pyurple	Regular Seller	kota surabaya	38.0	4.9	56.0	0.80	Jam
2607	KiosMeme	https://shopee.co.id/i2318	Regular Seller	kota surabaya	76.0	4.8	89.0	0.96	Jam
2608	sucindah6	https://shopee.co.id/sucindah6	Regular Seller	kota surabaya	46.0	4.9	66.0	0.94	Hari

2609 rows x 19 columns

Gambar 6.18 Pratinjau Data UMKM Shopee Hasil *Preprocess*.

Dilakukan analisis terhadap data hasil *preprocess*, sehingga didapatkan fakta-fakta seperti berikut.

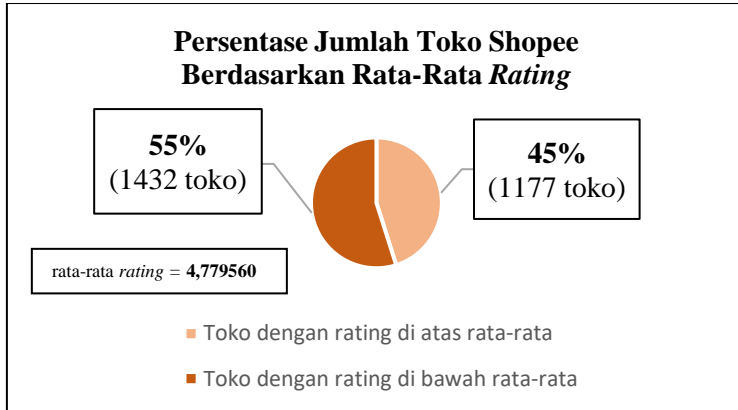
- Jumlah UMKM terbanyak berdasarkan tipe penjual adalah **2207** oleh *Regular Seller*, sedangkan paling sedikit sebanyak **6** oleh *Shopee Mall*. Persentase jumlah UMKM makanan dan minuman Shopee berdasarkan tipe penjual dapat dilihat pada Gambar 6.19.



Gambar 6.19 Perbandingan Data UMKM Shopee Berdasarkan Tipe *Seller*.

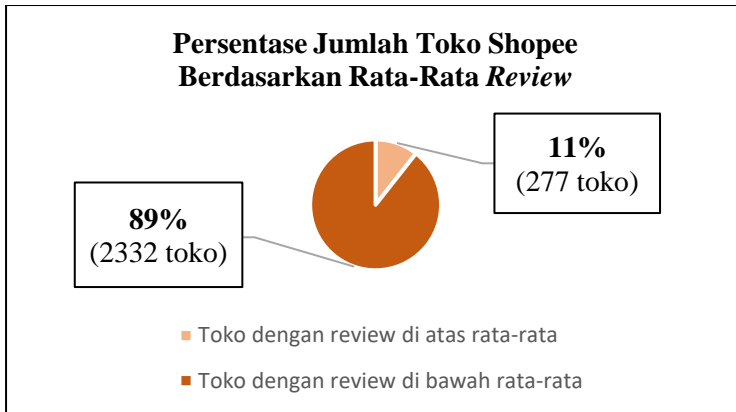
- b) Rata-rata jumlah *follower* UMKM makanan dan minuman di Shopee adalah **1612** dengan **99** UMKM tidak memiliki *review* sama sekali, dan sebanyak **271** UMKM memiliki jumlah *follower* di atas rata-rata.

Dari **2.609** UMKM didapatkan rata-rata *rating* sebesar **4,779560**. Terdapat 1177 UMKM telah memiliki *rating* di atas rata-rata. Persentase perbandingan UMKM dengan *rating* di atas dan di bawah rata-rata dapat dilihat pada Gambar 6.20. Dari data tersebut diketahui bahwa banyak UMKM makanan dan minuman di Surabaya pada Shopee yang memiliki *rating* cukup tinggi.



Gambar 6.20 Perbandingan Data UMKM Shopee Berdasarkan Rata-Rata *Rating*.

- c) Rata-rata jumlah *review* yang didapatkan UMKM adalah sebanyak **2.468**. Terdapat sebanyak **484** UMKM belum mendapatkan *review* sama sekali. Sebanyak **277** UMKM memiliki *review* di atas rata-rata atau sekitar **11%** dari 2609 UMKM. Hal tersebut masih menandakan terdapat kesenjangan jumlah *review* yang dimiliki oleh UMKM. Grafik persentase jumlah toko Shopee berdasarkan rata-rata *review* dapat dilihat pada Gambar 6.21.



Gambar 6.21 Perbandingan Data UMKM Shopee Berdasarkan Rata-Rata *Review*.

6.1.3 Google My Business

Data UMKM pada Google My Business merupakan data UMKM yang memiliki lokasi secara *offline* atau lokal. Dari UMKM-UMKM lokal tersebut didapatkan atribut-atribut yang disimpan dalam Google My Business. Dari hasil *scraping*, didapatkan data UMKM sebanyak **310.367**. Pratinjau Data UMKM Google My Business dapat dilihat pada Gambar 6.22.

	name	website	rating	numOf_google_review	price	category	address	operational_hours	phone	instagra
0	Cak Topa Degan Menur	NaN	NaN		NaN	Afghani restaurant	Jl. Pucang Jajar Tengah Jl. Pucang Anom No.71...	NaN	NaN	Na
1	HARRIS Hotel & Conventions Gubeng - Surabaya	https://booking.tauzia.com/54HARRIS-Gubeng?ul...	4.5	7,077 Google reviews	NaN	4-star hotel	Jl. Bangka No.05-15 Gubeng Kec. Gubeng Kota...	NaN	+62 31 5011100	Na
2	MaxOne Hotel Dharmahasuda	https://www.maxonehotels.com/dharmahasuda	4.4	2,060 Google reviews	NaN	3-star hotel	Dharmahasuda No.109, Mojo, Kec. Gubeng Ko...	NaN	+62 31 5957258	Na
3	Yaman Kuliner	NaN	4.3	1,215 Google reviews	\$\$	Middle Eastern restaurant	Jl. KH Maa Mansyur No.92, Ampel, Kec. Senampur...	Thursday 9AM–10PM Friday 9AM–10PM Saturday 9...	+62 812-4617-2422	Na
4	RedDoorz near Gubeng Station	https://www.reddoorz.com/id-hotel/inidonesia...	3.5	590 Google reviews	NaN	3-star hotel	Jl. Jawa No.22 Gubeng, Kec. Gubeng Kota SEY...	NaN	+62 815-3821-1110	Na

Gambar 6.22 Pratinjau Data UMKM Google My Business Hasil *Scraping*.

Namun, dari hasil *scraping* tersebut masih terdapat banyak data yang tidak relevan sehingga dilakukan *preprocess*. *Preprocess* yang dilakukan untuk mendapatkan data yang relevan di

antaranya adalah menghapus UMKM dengan kaegori yang tidak relevan dengan makanan dan minuman, dan menghapus duplikasi UMKM berdasarkan *name*, *website*, dan *adres*. Terdapat beberapa data *null* atau *NaN* yang didapatkan karena adanya kemungkinan terdapat masalah saat pengambilan data, seperti tidak *load* halaman *website* dengan sempurna. Contohnya adalah sebanyak **131** dari **310.367** data hasil *scraping* Google My Business tidak memiliki nama, padahal seharusnya semua UMKM memiliki nama, sehingga data tersebut harus dihapus. Didapatkan data UMKM sebanyak **39.347** di Surabaya yang dari hasil *preprocess*. Pratinjau data UMKM Google My Business hasil *preprocess* dapat dilihat pada Gambar 6.23. Keseluruhan data UMKM makanan dan minuman Google My Business dapat dilihat pada LAMPIRAN Tabel B.11.

	name	website	rating	numOf_google_review	price	category	address	kecamatan	kelurahan	operational_hours	operational_hours_duration
0	cak lopa degan menu	website not found	NaN	NaN	price not found	afghani restaurant	jl. pucang jajar tengah jl. pucang anom no.71...	gubeng	gubeng	NaN	0.0
2	depot lyn	website not found	4.5	2.0	price not found	afghani restaurant	jl. galaxi klampis asri barat i no.9, semolow...	sukolilo	semolowaru	Thursday 8AM– 8PM,Friday 8AM– 8PM,Saturday 8AM...	84.0
3	mis kivism	website not found	NaN	NaN	price not found	afghani restaurant	jl. simpang damo permai utara no.25, lontar...	wonokromo	damo	NaN	0.0
4	kolampancing bandeng cak khan	website not found	2.5	2.0	price not found	afghani restaurant	osowilangon30, benovo, surabaya city, east jav...	pakal	benovo	Thursday 8AM– 6PM,Friday 8AM– 6PM,Saturday 8AM...	70.0
5	martabak hawaii rungkut	website not found	4.5	96.0	price not found	martabak restaurant	jalan raya kali rungkut lor. kali rungkut. kec...	rungkut	kali rungkut	Thursday 12– 11PM,Friday 1– 11PM,Saturday 12–1...	76.0

Gambar 6.23 Pratinjau Data UMKM Google My Business Hasil *Preprocess*.

6.2 Integrasi Data

Proses integrasi data atau *data integration* menggunakan metode *entity resolution* dengan melihat nilai similaritas atau kemiripan dari suatu atribut data. Dipilih atribut-atribut data yang sama yaitu atribut *phone*, *website*, *instagram* dan *name*. Terdapat 2 hasil data integrasi dengan tujuan untuk melihat UMKM, yaitu secara *online* dengan mengintegrasikan data UMKM pada Tokopedia dengan data UMKM pada Shopee, dan integrasi UMKM lokal/*online* (memiliki lokasi lokal dan memiliki atau tidak memiliki UMKM *online*) dengan

mengintegrasikan data UMKM Google My Business, Tokopedia, dan Shopee.

6.2.1 Tokopedia, dan Shopee

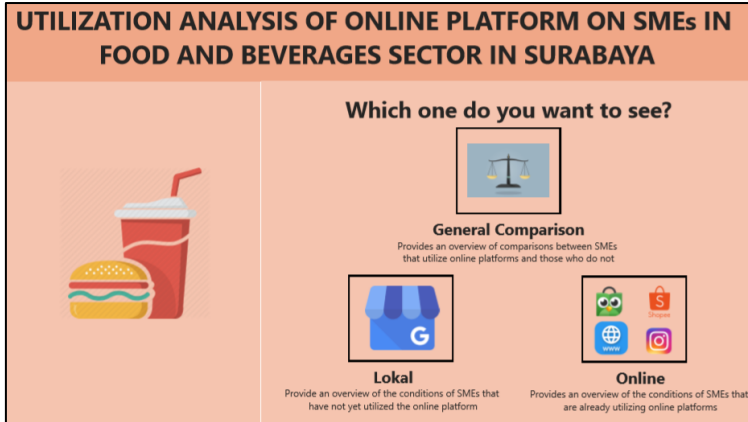
Integrasi pertama untuk melihat UMKM *online* dengan mengintegrasikan data Tokopedia dan Shopee, serta memanfaatkan hasil tes *similarity* dengan hasil *scraping* Instagram dari masing-masing platform. Hasil lengkap data bisa dilihat pada LAMPIRAN Tabel B.12.

6.2.2 Google My Business, Tokopedia, dan Shopee

Integrasi kedua dilakukan terhadap data Google My Business, Tokopedia dan Shopee untuk melihat UMKM lokal serta lokal & *online*. Hasil lengkap data bisa dilihat pada LAMPIRAN Tabel B.13.

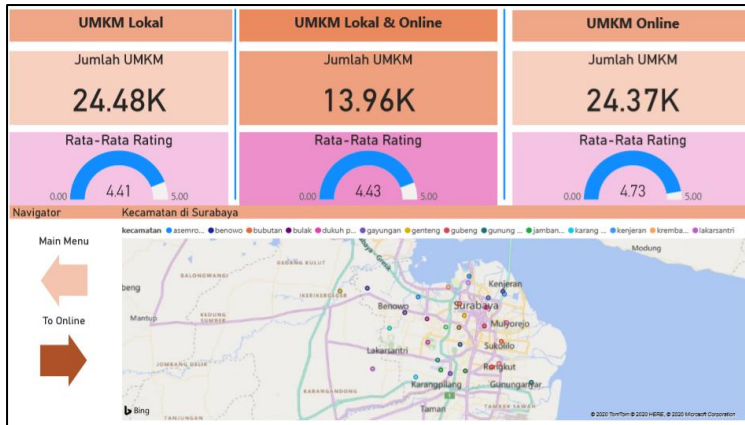
6.3 Visualisasi dan Analisis

Visualisasi informasi dibedakan menjadi tiga kategori, yaitu *general comparison*, *offline*, dan *online*. *General comparison* menampilkan visualisasi informasi mengenai perbandingan antara UMKM *offline*, *online*, dan *online/offline*. Pada menu *offline* akan menampilkan informasi mengenai UMKM-UMKM lokal. Pada halaman *online* menampilkan informasi mengenai UMKM-UMKM yang berada di pasar *online* (Tokopedia dan Shopee). Tampilan laman visualisasi dapat dilihat pada Gambar 6.24.



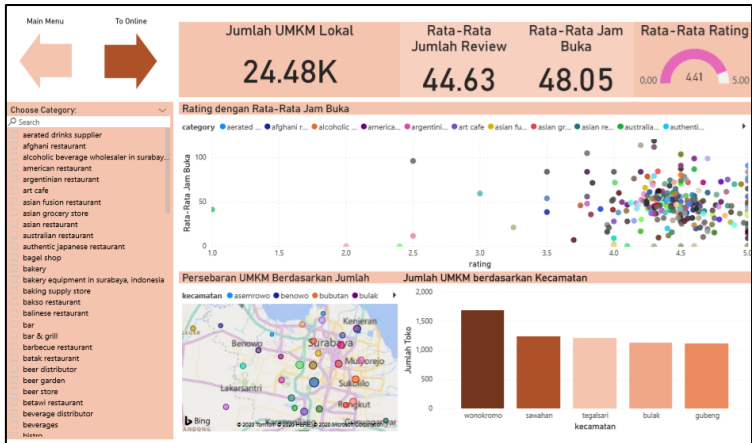
Gambar 6.24 Laman Visualisasi Informasi.

Pada halaman *general comparison* ditampilkan informasi mengenai jumlah UMKM lokal, *online*, serta lokal & *online*. Dari visualisasi yang ada ditunjukkan bahwa pelanggan lebih cenderung memberikan *rating* lebih tinggi dibandingkan dengan UMKM lokal dilihat dari rata-rata *rating* antara UMKM lokal dan *online*. Ditampilkan juga informasi mengenai jumlah UMKM lokal yang telah menggunakan platform daring (*website*, Instagram, Tokopedia dan Shopee) berjumlah **13.960** atau **35%** dari **39.347** jumlah keseluruhan UMKM dilihat dari data UMKM Google My Business. Visualisasi *general comparison* dapat dilihat pada Gambar 6.25.



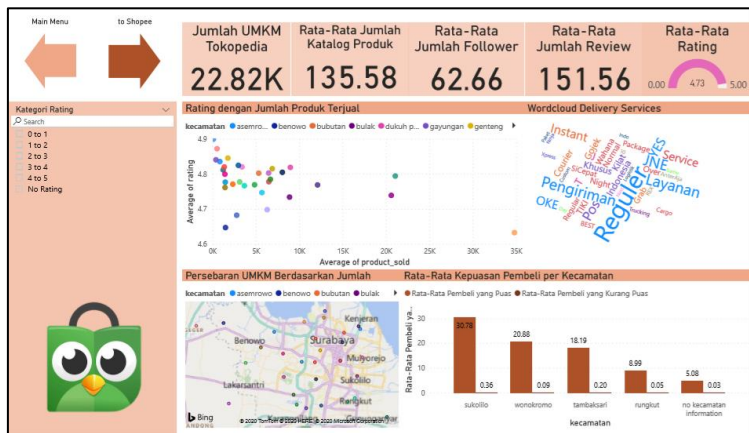
Gambar 6.25 Halaman *General Comparison*.

Pada halaman *UMKM Offline*, ditampilkan jumlah UMKM lokal (tanpa *online*), rata-rata jumlah *review*, rata-rata jam operasional, rata-rata *rating*, korelasi rata-rata jam operaional dan *rating*, serta persebaran UMKM berdasarkan kecamatan. Didapatkan informasi bahwa rata-rata jam operasional tidak menjamin tinggi rendahnya *rating* yang diberikan oleh pelanggan karena banyak yang jam operasionalnya rendah tetapi mendapatkan *rating* cukup tinggi. Persebaran UMKM *offline* makanan dan minuman banyak terdapat di kecamatan Wonokromo, Sawahan, Tegalsari, Bulak, dan Gubeng. Kelima daerah tersebut berada di antara area Surabaya Selatan atau Surabaya Pusat, sehingga didapatkan pusat persebaran UMKM *offline* makanan dan minuman berada pada dua area tersebut. Visualisasi pada halaman *UMKM offline* dapat dilihat pada Gambar 6.26.

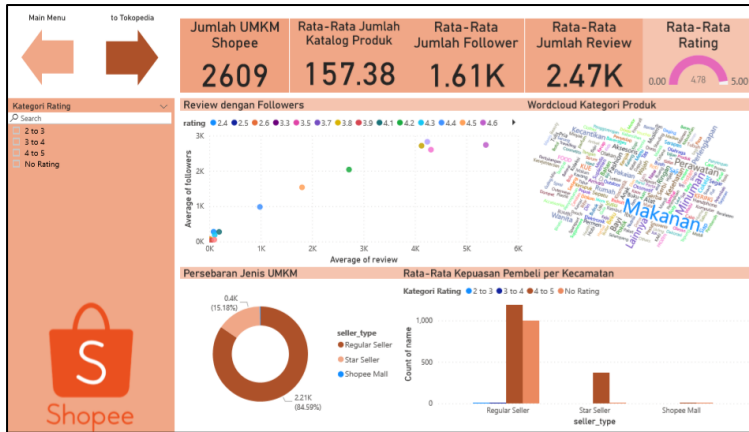


Gambar 6.26 Halaman Visualisasi Data UMKM Lokal.

Ditampilkan kondisi UMKM *online* pada platform Tokopedia dan Shopee seperti pada **Gambar 6.27**, dan **Gambar 6.28**. Dari kedua halaman tersebut didapatkan bahwa terdapat korelasi atau hubungan antara variabel rata-rata jumlah *follower* dengan rata-rata jumlah *review*. Semakin banyak rata-rata jumlah *follower*, maka semakin banyak rata-rata jumlah *review*.



Gambar 6.27 Halaman Visualisasi Data UMKM Tokopedia.



Gambar 6.28 Halaman Visualisasi Data UMKM Shopee.

BAB VII PENUTUP

Bab ini membahas kesimpulan penelitian yang telah dilakukan dan saran yang diusulkan. Kesimpulan ini diharapkan dapat menjawab rumusan masalah penelitian dan saran untuk penelitian serupa di masa mendatang.

7.1 Kesimpulan

Berdasarkan pengumpulan, pengolahan, integrasi, serta visualisasi dan analisis data, diperoleh beberapa kesimpulan sebagai berikut.

1. Data UMKM tersedia secara daring dalam format beragam dan berlimpah. Beberapa data (*golden data*) menjadi penting dan bermakna untuk dianalisis. Pada tugas akhir ini tujuan analisis adalah untuk mengetahui kondisi UMKM dalam penggunaan platform daring yang dibagi menjadi pasar *online (marketplace)*, direktori bisnis lokal (Google My Business), dan sosial media (Instagram). Dilakukan observasi terhadap pasar *online* (halaman pencarian, kategori produk, detail produk, dan detail UMKM), direktori bisnis lokal (halaman pencarian, dan detail UMKM), dan sosial media (halaman pencarian). Dalam melakukan observasi ditemukan tantangan untuk menentukan lokasi data yang unik. Pencarian keunikan lokasi data pada halaman *website* dapat dibantu dengan menggunakan fitur penentuan lokasi data pada *browser*. Dari hasil observasi dirumuskan skema data untuk menampung data sementara sebagai berikut:
 - a. Pada pasar *online* didapatkan skema data yang terbagi menjadi data kategori produk, informasi produk, dan informasi UMKM.
 - b. Pada direktori bisnis lokal didapatkan skema data untuk menyimpan informasi UMKM.
 - c. Pada sosial media didapatkan skema data untuk menyimpan hasil pencarian UMKM.

Skema data yang dihasilkan dapat menjadi referensi atau

gambaran kondisi data UMKM pada platform daring untuk penelitian berikutnya. Sehingga, pada penelitian berikutnya dapat dilakukan pengembangan dan penyesuaian skema data sesuai dengan kondisi yang relevan, misal perubahan data yang disajikan pada platform daring atau target data yang diambil.

2. Pembuatan kode *scraping* digunakan untuk pengambilan data dari *website* secara otomatis. Terdapat dua teknik yang digunakan, yaitu *raw* dan *headless*. Pengambilan data secara *raw* memiliki keunggulan dalam kompatibilitas perangkat Windows, namun lebih lambat (lebih dari 7 hari pada direktori bisnis lokal). Berbeda dengan pengambilan data secara *headless*, lebih kompatibel dengan perangkat Linux, namun dapat berjalan lebih cepat (4 hari pada direktori bisnis lokal). Maka, pengambilan data secara *raw* lebih cocok digunakan pada perangkat berbasis Windows, dan *headless* lebih cocok digunakan pada perangkat berbasis Linux. Tantangan dalam pengembangan kode *scraping* di antaranya mengenai pembaruan kode program yang harus selalu sesuai dengan pembaruan *website* sehingga memberikan peringatan ketika terjadi perubahan pada halaman yang diambil datanya, serta tidak hanya berisi pengambilan data tetapi juga dapat mengakomodir perilaku-perilaku tiap *website* (*close pop-up*, klik tombol tertentu) yang perlu dilakukan pada saat proses pengambilan data. Kode *scraping* yang telah dibuat dapat menjadi bahan referensi atau *template* untuk penelitian berikutnya sebagai sebuah cara untuk mengambil data melalui *web* sesuai dengan kondisi yang relevan, misal pembaruan *website* atau perubahan struktur halaman *website* terkait lokasi penyimpanan data.
3. Pengolahan data dengan menstrukturkan, dan menyeragamkan format data agar dapat dilakukan analisis terhadap data, seperti membandingkan, menjumlahkan, dan agregasi. Integrasi dilakukan untuk menggabungkan data sehingga dapat dilakukan analisis antar platform dengan melihat kemiripan makna atribut. Dari proses penggabungan data dihasilkan skema data sebagai berikut:

- a. Integrasi pasar *online* yang terdiri dari gabungan atribut dari masing-masing pasar *online*, dan
- b. Integrasi data direktori bisnis lokal dengan pasar *online* sehingga dihasilkan data gabungan dari kedua platform tersebut.

Dalam melakukan pengolahan dan integrasi data terdapat beberapa tantangan di antaranya adalah mengenai keseragaman data. Data dengan detil atau satuan berbeda dilakukan konversi hingga seragam, serta perlunya menentukan *threshold* integrasi yang optimal dengan melihat nilai presisi. Nilai presisi paling tinggi yang didapatkan adalah 0,69 dengan *threshold* 0,8 untuk kategori makanan dan minuman. *Threshold* dapat menjadi masukan untuk penelitian berikutnya sebagai angka optimal dalam melakukan integrasi data.

4. Visualisasi informasi dilakukan untuk dapat melihat kondisi UMKM makanan dan minuman di Surabaya berdasarkan hasil data yang telah dikumpulkan melalui *scraping*. Pemerintah Kota Surabaya maupun peneliti dapat melihat visualisasi yang dihasilkan untuk melihat informasi yang disajikan dari data yang dikumpulkan sebagai bentuk ketersediaan informasi bisnis dan kondisi pemanfaatan platform daring oleh UMKM makanan dan minuman di Surabaya. Visualisasi dilakukan dengan melakukan agregasi atau pengelompokan sehingga didapatkan ringkasan data. Dari analisis visualisasi didapatkan hasil bahwa:
 - a. Tidak banyak UMKM makanan dan minuman lokal yang menggunakan platform daring, terlihat dari selisih jumlah UMKM lokal yang telah menggunakan platform daring dibanding UMKM lokal cukup jauh.
 - b. Platform daring memiliki dampak yang baik terhadap reputasi UMKM dengan melihat rata-rata *rating* UMKM lokal yang telah menggunakan platform daring lebih tinggi dibandingkan UMKM lokal.

- c. Selain itu, dari UMKM yang telah menggunakan platform daring penggunaannya masih belum optimal, karena terlihat dari proporsi penggunaan fitur platform daring seperti *star seller* masih lebih sedikit dibandingkan dengan *seller* biasa pada platform Shopee.

7.2 Saran

Dalam pengerjaan tugas akhir, terdapat beberapa saran yang diharapkan dapat bermanfaat bagi pihak terkait maupun untuk pengembangan penelitian ke depan yang dijelaskan sebagai berikut.

1. Bagi penelitian selanjutnya
 - a. Perlu mengambil data-data mengenai produk, seperti nama-nama, jumlah penjualan, dan *rating* tiap produk yang dijual, sehingga dapat dilakukan analisis terhadap kondisi toko pada platform daring berdasarkan produk yang dijual.
 - b. Perlu adanya tambahan atribut untuk melakukan integrasi dengan metode *similarity*, seperti *image processing* dengan melihat kemiripan gambar profil toko atau gambar produk pada toko, karena didapatkan bahwa banyak toko yang sama memiliki gambar profil atau produk yang sama.
 - c. Menskenariokan pengujian kualitas data hasil integrasi dengan menggunakan *threshold* yang berbeda-beda tiap hasil integrasi yang dilakukan, kemudian melakukan pemilihan *threshold* optimal dengan menghitung jumlah kemungkinan data yang benar diintegrasikan menggunakan penghitungan nilai presisi dikali jumlah data yang berhasil diintegrasikan.

2. Bagi Pemerintah kota Surabaya

Analisis yang dilakukan adalah analisis deskriptif, sehingga menampilkan kondisi eksisting. Dalam mengetahui kondisi yang akan terjadi dan apa yang perlu dilakukan ke depannya perlu dilakukan analisis prediktif dan preskriptif, tetapi masih

tidak memungkinkan menggunakan data saat ini. Dengan menambah data-data yang terkait dengan UMKM yang dimiliki oleh Pemerintah Kota Surabaya, misal hasil survei usaha, analisis prediktif dan preskriptif menjadi dapat dilakukan. Sehingga, analisis pendukung yang lebih menyuruh dapat dilakukan untuk membantu pengambilan keputusan terkait pengembangan UMKM makanan dan minuman di Surabaya.

Halaman ini sengaja dikosongkan.

DAFTAR PUSTAKA

- [1] Bank Indonesia dan LPPI, “Profil Bisnis Usaha Mikro, Kecil dan Menengah (UMKM),” *Bank Indones. dan LPPI*, pp. 18–20, 2015.
- [2] W. P. Utomo, “Indonesia Millennial Report 2019,” 2019. [Online]. Available: <https://www.idntimes.com/indonesiamillennialreport2019>.
- [3] B. P. Statistik, *Statistik E-Commerce 2019*. 2019.
- [4] R. B. Mahardhika, *MENGENAL INDUSTRI MAKANAN DAN MINUMAN DI ERA INDUSTRI 4.0*. Yogyakarta: Forbil Institute.
- [5] C. E. Carpio and K. Y. Lange, “Trends in e-commerce for the food marketing system,” *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.*, vol. 10, no. January, 2015, doi: 10.1079/PAVSNNR201510023.
- [6] H. G. Cheng and M. R. Phillips, “Secondary analysis of existing data: opportunities and implementation,” *Shanghai Arch. Psychiatry*, vol. 26, no. 6, pp. 371–375, 2014, doi: 10.11919/j.issn.1002-0829.214171.
- [7] C. Wang, K. Chen, T. Bo, J. Pei, V. S. Tseng, and R. Goebel, *Trends and Applications in Knowledge Discovery*, no. April. 2013.
- [8] V. Draxl, “Web Scraping Data Extraction from websites,” 2018.
- [9] R. Shah, K. Pathan, A. Masurkar, S. Rewatkar, and N. Vengurlekar, “Comparison of E-commerce Products using web mining,” *Int. J. Sci. Res. Publ.*, vol. 6, no. 5, p. 640, 2016, [Online]. Available: www.ijsrp.org.
- [10] J. Auer and I. Boettcher, “From price collection to price data analytics How new large data sources require price statisticians to re-think their index compilation procedures . Experiences from web-scraped and scanner

- data,” no. May 2016, pp. 1–26.
- [11] R. S. Chaulagain, S. Pandey, S. R. Basnet, and S. Shakya, “Cloud Based Web Scraping for Big Data Applications,” *Proc. - 2nd IEEE Int. Conf. Smart Cloud, SmartCloud 2017*, no. August 2018, pp. 138–143, 2017, doi: 10.1109/SmartCloud.2017.28.
 - [12] Badan Kebijakan and P. P. R. Fiskal, “Kebijakan Antisipasi Krisis Tahun 2012 Melalui Program Kredit Usaha Rakyat (KUR),” *Badan Kebijak. Fiskal*, p. 47, 2012.
 - [13] M. E. Megel and J. A. Heermann, “Methods of data collection,” *Plast. Surg. Nurs.*, vol. 14, no. 2, pp. 109–110, 1994, doi: 10.5005/jp/books/13075_10.
 - [14] R. Mitchell, *Web Scraping with Python*, First Edit. Sebastopol: O’Reilly, 316AD.
 - [15] V. Gour, S. S. Sarangdevot, G. S. Tanwar, and A. Sharma, “Improve Performance of Extract, Transform and Load ({ETL}) in Data Warehouse,” *Int. J. Comput. Sci. Eng.*, vol. 1, no. 3, pp. 786–789, 2010.
 - [16] Pemerintahan Kota Surabaya, “Bagian Administrasi Pemerintah dan Otonomi Daerah | Kota Surabaya,” 2018.
https://pemerintahan.surabaya.go.id/home/kecamatan_kelurahan (accessed Apr. 30, 2020).
 - [17] S. Faisal, “ANALISIS PEMANFAATAN PLATFORM ONLINE DALAM PROSES PENJUALAN PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA.” 2020.
 - [18] R. V. Krejcie and D. W. Morgan, “DETERMINING SAMPLE SIZE FOR RESEARCH ACTIVITIES,” *Educ. Psychol. Meas.*, vol. 30, pp. 607–610, 1970, doi: 10.1261/rna.2763111.
 - [19] A. Ramadhan, “ANALISIS PEMANFAATAN

PLATFORM DARING PADA UMKM BIDANG KESEHATAN DI SURABAYA.” 2020.

Halaman ini sengaja dikosongkan.

LAMPIRAN A. KODE PROGRAM

Pada bagian ini akan diberikan kode *scraping* pada masing-masing platform.

A. Kode *Scraping* Data

Pada bagian ini dilampirkan kode yang digunakan untuk *scraping* data pada tiap platform.

Tabel A.1 Kode Program *Scraping* Kategori Tokopedia.

Kode <i>scraping</i> kategori makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/1Jweca6_UARbqMxNuGoxSM_zCV82nDsea/view?usp=sharing

Tabel A.2 Kode Program *Scraping* Produk Tokopedia.

Kode <i>scraping</i> produk makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/1LKoidnuOKPYs8K9wIiSeNiK9j3NvufNZ/view?usp=sharing

Tabel A.3 Kode Program *Preprocess URL* Produk Menjadi *URL* UMKM Tokopedia.

Kode <i>preprocess URL</i> produk menjadi <i>URL</i> UMKM	
URL	https://drive.google.com/file/d/1jQIvDbzVpFkczybAX5vtu7CIoHoaiqXG/view?usp=sharing

Tabel A.4 Kode Program *Scraping* Informasi UMKM Makanan dan Minuman Tokopedia.

Kode <i>scraping</i> data UMKM makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/1YwTERkx4YyMfbGq45jnyz22nAvB3XhNS/view?usp=sharing

Tabel A.5 Kode Program *Scraping* Kategori Makanan dan Minuman Shopee.

Kode <i>scraping</i> data kategori makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1XWyDaN2zg4QW6WVkcSY2S-KAghZTuRb/view?usp=sharing

Tabel A.6 Kode Program *Scraping* Produk Makanan dan Minuman Shopee.

Kode <i>scraping</i> data produk makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1dTye3HbHhSsdrgfb1QH7RUsvLPd693Lk/view?usp=sharing

Tabel A.7 Kode Program *Scraping URL* UMKM Makanan dan Minuman Shopee.

Kode <i>scraping</i> data <i>URL</i> UMKM makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1UImCfr7jyIiKlO5K7hCJ8h5WUbcEw9w9/view?usp=sharing

Tabel A.8 Kode Program *Scraping* Informasi UMKM Makanan dan Minuman Shopee.

Kode <i>scraping</i> data informasi UMKM makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1DK3tlry83v2tkmFXZkBp4XQOULLQTj7/view?usp=sharing

Tabel A.9 Kode Program *Scraping* Informasi UMKM Makanan dan Minuman Google My Business.

Kode <i>scraping</i> data informasi UMKM makanan dan minuman di Google My Business. Dibagi menjadi lima bagian (<i>batch</i>) untuk mempercepat proses <i>scraping</i> dan dijalankan secara bersamaan pada VPS.	
URL part 1	https://drive.google.com/file/d/1DK3tlry83v2tkmFX_ZkBp4XQOULLQTj7/view?usp=sharing
URL part 2	https://drive.google.com/file/d/11VOWIBb0Eimki8-f_KIP_MH3_wrHVKUR/view?usp=sharing
URL part 3	https://drive.google.com/file/d/1WyKbKMvwOhz4Tr4Pms6vhpO_LgnREhRB/view?usp=sharing
URL part 4	https://drive.google.com/file/d/18AzDKIG3aLbNX4EzZgqaNhb8IcKonVGB/view?usp=sharing
URL part 5	https://drive.google.com/file/d/1OoFQs6tzFeTLGpj8yObj93hfMGt_sKqk/view?usp=sharing

B. Kode *Preprocess* Data UMKM

Pada bagian ini dilampirkan kode yang digunakan untuk *preprocess* data informasi UMKM pada tiap platform.

Tabel A.10 Kode Program *Preprocess* Data UMKM Makanan dan Minuman Tokopedia.

Kode <i>preprocess</i> data UMKM makanan dan minuman di Tokopedia	
URL	https://drive.google.com/file/d/1ynQtlCxEx2jxLJLC_L9SldbbYujjwmvdLp/view?usp=sharing

Tabel A.11 Kode Program *Preprocess* Data UMKM Makanan dan Minuman Shopee.

Kode <i>preprocess</i> data UMKM makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1x29FnGlthhwpqNQ9oXv-NjZpta9OND8A/view?usp=sharing

Tabel A.12 Kode Program *Preprocess* Data UMKM Makanan dan Minuman Google My Business.

Kode <i>preprocess</i> data UMKM makanan dan minuman di Google My Business	
URL	https://drive.google.com/file/d/15FXT6bL6MT7EpLXPQx0U39qbWkZYxa6B/view?usp=sharing

C. Kode Integrasi Data UMKM

Pada bagian ini dilampirkan kode yang digunakan untuk integrasi data UMKM. Integrasi dibagi menjadi dua, yaitu integrasi data UMKM Tokopedia dengan Shopee, dan Google My Business dengan Tokopedia dan Shopee.

Tabel A.13 Kode Program Integrasi Data UMKM Makanan dan Minuman Tokopedia dengan Shopee.

Kode integrasi data UMKM antara dua platform, yaitu Tokopedia dan Shopee	
URL	https://drive.google.com/file/d/12ltsr-qKQUcU6t4FsEdIfFgyYmN6OQJ-/view?usp=sharing

Tabel A.14 Kode Program Integrasi Data UMKM Makanan dan Minuman Google My Business dengan Tokopedia dan Shopee.

Kode integrasi data UMKM antara tiga platform, yaitu Gooogle My Business dengan Tokopedia dan Shopee	
<i>URL</i>	https://drive.google.com/file/d/1fhDzol9USwywjaa1FED5r-u6peeBEevY/view?usp=sharing

LAMPIRAN B. DATA HASIL SCRAPING, PREPROCESS, DAN INTEGRASI

Pada bagian ini akan diberikan hasil *scraping* dan *preprocess* pada masing-masing platform.

D. Data Hasil Scraping

Tabel B.1 Data Kategori Makanan dan Minuman Tokopedia.

Data hasil <i>scraping</i> kategori makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/15kZmnXhV2Largd9slAEwp2tFNa44pkTy/view?usp=sharing

Tabel B.2 Data Produk Makanan dan Minuman Tokopedia.

Data hasil <i>scraping</i> produk makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/1PXCoe0ziNVM1ElGfu3JbOZSa-kn2GsDc/view?usp=sharing

Tabel B.3 Data URL UMKM Makanan dan Minuman Tokopedia.

Data URL UMKM hasil <i>preprocess</i> URL produk makanan dan minuman Tokopedia	
URL	https://drive.google.com/file/d/1swDPGnIrU8uDjRBUf6bWXezuZjaorn5T/view?usp=sharing

Tabel B.4 Data UMKM Makanan dan Minuman Tokopedia.

Data UMKM makanan dan minuman hasil <i>scraping</i> di Tokopedia	
URL	https://drive.google.com/file/d/1CyVLKxwKmmNajmxEApPuVc71oNtHXFCL/view?usp=sharing

Tabel B.5 Data Kategori Makanan dan Minuman Shopee.

Data kategori makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1FBnmmcaBzDmWedQVLIQjv_vkOBJhAuN4/view?usp=sharing

Tabel B.6 Data Produk Makanan dan Minuman Shopee.

Data produk makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/1j0wocCki2QISh3QLq7AtXq-9z47r-oyf/view?usp=sharing

Tabel B.7 Data URL UMKM Makanan dan Minuman Shopee.

Data URL UMKM makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/16MqUVQpmCtyZD_N-oC6WW-2RY2OW8o6f/view?usp=sharing

Tabel B.8 Data UMKM Makanan dan Minuman Shopee.

Data UMKM makanan dan minuman di Shopee	
URL	https://drive.google.com/file/d/10xXOtbccfSqsqt4yVR5Jn0wIPw1rw_3/view?usp=sharing

E. Data UMKM Hasil *Preprocess*

Tabel B.9 Data UMKM Makanan dan Minuman Tokopedia Hasil *Preprocess*.

Data UMKM makanan dan minuman di Tokopedia hasil <i>preprocess</i>	
URL	https://drive.google.com/file/d/1jB_BBIUE9ON6bIFQb7cfaYzorvj57beQ/view?usp=sharing

Tabel B.10 Data UMKM Makanan dan Minuman Shopee Hasil *Preprocess*.

Data UMKM makanan dan minuman di Shopee hasil <i>preprocess</i>	
URL	https://drive.google.com/file/d/1eofy2N0jzvRB1TN1hmEbkyap_T2ohSJ1/view?usp=sharing

Tabel B.11 Data UMKM Makanan dan Minuman Google My Business Hasil *Preprocess*.

Data UMKM makanan dan minuman di Shopee hasil <i>preprocess</i>	
URL	https://drive.google.com/file/d/1KUBX4fX_6QQtmzA2LoPYrasIxthqiOKT/view?usp=sharing

F. Data UMKM Hasil Integrasi

Data UMKM hasil integrasi dibagi menjadi dua, yaitu integrasi data UMKM Tokopedia dengan Shopee, dan Google My Business dengan Tokopedia dan Shopee.

Tabel B.12 Data UMKM Makanan dan Minuman Hasil Integrasi Tokopedia dengan Shopee.

Data UMKM makanan dan minuman hasil integrasi antara dua platform, yaitu Tokopedia dengan Shopee	
URL	https://drive.google.com/file/d/1qYv40VPs1Yzujfv3Tlf_6c3KhA5EI4D/view?usp=sharing

Tabel B.13 Data UMKM Makanan dan Minuman Hasil Integrasi Google My Business dengan Tokopedia dan Shopee.

Data UMKM makanan dan minuman hasil integrasi antara tiga platform, yaitu Google My Business dengan Tokopedia dan Shopee	
URL	https://drive.google.com/file/d/1cc00Ks-SYRLQxqRIH1jm4j_CGBWHpG8S/view?usp=sharing

Halaman ini sengaja dikosongkan.

LAMPIRAN C. *TABEL CHI SQUARE*

Tabel C.1 Tabel *Chi Square*.

DF	Probabilitas			
	0.5	0.1	0.05	0.01
1	0.45494	2.70554	3.84146	6.63490
2	1.38629	4.60517	5.99146	9.21034
3	2.36597	6.25139	7.81473	11.34487
4	3.35669	7.77944	9.48773	13.27670
5	4.35146	9.23636	11.07050	15.08627
6	5.34812	10.64464	12.59159	16.81189
7	6.34581	12.01704	14.06714	18.47531
8	7.34412	13.36157	15.50731	20.09024
9	8.34283	14.68366	16.91898	21.66599
10	9.34182	15.98718	18.30704	23.20925
11	10.34100	17.27501	19.67514	24.72497
12	11.34032	18.54935	21.02607	26.21697
13	12.33976	19.81193	22.36203	27.68825
14	13.33927	21.06414	23.68479	29.14124
15	14.33886	22.30713	24.99579	30.57791

Halaman ini sengaja dikosongkan.

BIODATA PENULIS



Penulis lahir di Sidoarjo pada tanggal 23 Maret 1998 dan merupakan anak pertama dari dua bersaudara pasangan Bapak Sugeng Sutopo dan Ibu Endang Wahyu Setyowati. Penulis menempuh Pendidikan formal di SD Negeri Jimbaran Kulon, SMP Negeri 1 Wonoayu dan SMA Negeri 3 Sidoarjo.

Pada tahun 2016 penulis melanjutkan pendidikan jenjang sarjana dengan jalur SNMPTN di Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi (FTIK) yang sekarang menjadi Fakultas Teknologi Elektro dan Informatika Cerdas (FTEIC) - Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 05211640000023. Selama masa perkuliahan penulis aktif mengikuti organisasi mahasiswa, seperti Himpunan Mahasiswa Sistem Informasi dengan jabatan *staff* hingga kepala departemen bidang Riset dan Aplikasi Teknologi. Penulis juga aktif dalam kegiatan kepanitiaan, keilmiahan, dan perlombaan, seperti ISE, GEMASTIK, APPCELERATE, SHELL LIVEWIRE, dan lainnya.

Pada tahun keempat, penulis memilih untuk fokus di bidang Sistem *Enterprise* khususnya Bisnis Digital. Oleh karena itu, penulis terdaftar sebagai mahasiswa tugas akhir di Laboratorium Sistem *Enterprise*, Departemen Sistem Informasi, ITS. Penulis dapat dihubungi melalui email ubaiyusuf@gmail.com.