



ITS

Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - IS184853

**ANALISIS PEMANFAATAN PLATFORM ONLINE
UNTUK MENGGAMBARKAN KETERSEDIAAN
INFORMASI BISNIS PADA UMKM KATEGORI FASHION
MUSLIM DI SURABAYA**

***UTILIZATION ANALYSIS OF ONLINE PLATFORM TO
DESCRIBE THE AVAILABILITY OF BUSINESS
INFORMATION IN SMEs MUSLIM FASHION IN
SURABAYA***

**SYAFRIE DWI FAISAL
0521 16 4000 0134**

**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 ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA



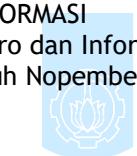
SYAFRIE DWI FAISAL
0521 16 4000 0134



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



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





ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh

ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh

UNDERGRADUATE THESIS - IS184853

UTILIZATION ANALYSIS OF ONLINE PLATFORM TO DESCRIBE THE AVAILABILITY OF BUSINESS INFORMATION IN SMEs MUSLIM FASHION IN SURABAYA

SYAFRIE DWI FAISAL
0521 16 4000 0134

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



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh

LEMBAR PENGESAHAN

ANALISIS PEMANFAATAN PLATFORM ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA

TUGAS AKHIR

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

Oleh:

SYAFRIE DWI FAISAL

NRP. 05211640000134

Surabaya, 13 Agustus 2020

**KEPALA
DEPARTEMEN SISTEM INFORMASI**

Dr. Mudja'hidin, S.T., M.T
NIP. 19701010 200312 1 001





LEMBAR PERSETUJUAN

ANALISIS PEMANFAATAN PLATFORM ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM 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:

SYAFRIE DWI FAISAL

NRP. 05211640000134



Disetujui Tim Penguji : Tanggal Ujian: 03 Juni 2020
Periode Wisuda : September 2020

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

Mahendrawati ER, ST, M.Sc, Ph.D

(Pembimbing I)

(Penguji I)

Erma Suryani, S.T., M.T., Ph.D

(Penguji II)

(Penguji II)



ANALISIS PEMANFAATAN PLATFORM ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA

Nama Mahasiswa : Syafrie Dwi Faisal
NRP : 0521 16 4000 0134
Departemen : Sistem Informasi FTIK-ITS
**Pembimbing I : Rully Agus Hendrawan, S.Kom,
M.Eng**

ABSTRAK

Saat ini banyak UMKM yang telah memanfaatkan platform online. UMKM memerlukan pengolahan dan analisis data untuk terus update dalam mengetahui keadaan bisnis.

Tidak adanya alat untuk mengumpulkan data dengan cepat dan minimnya pengetahuan dalam melakukan pengolahan data membuat UMKM dalam penggunaan platform online tidak dapat berjalan dengan maksimal. Padahal Pengolahan data yang benar akan menjadi sebuah informasi yang dapat berguna untuk menentukan keputusan dan mengetahui keadaan bisnis saat ini secara real-time. Industri yang akan dipilih adalah industri fashion muslim.

Terlebih dahulu dilakukan data scrapping untuk mengambil data platform online tokopedia, shopee, google business dan instagram. Hasil data kemudian dilakukan preprocessing agar data menjadi normal. Data dilakukan integrasi agar bisa dilihat

secara online, offline maupun online/offline. Dilanjutkan dengan penilaian kebenaran hasil pencocokan data dengan membandingkan hasil integrasi dan memilih hasil penilaian dengan rata-rata skor precision tertinggi. Kemudian dilakukan analisis secara deskriptif dengan melakukann visualisasi informasi.

Hasil dari penelitian ini berupa skema data, kumpulan data UMKM fashion muslim pada beberapa platform online, alur penggerjaan uji validitas hasil intergasi untuk menilai kehandalan data dan visualisasi informasi untuk analisis data. Hasil visulisasi informasi membuktikan bahwa secara kuantitas UMKM sudah baik dalam memanfaatkan platform online. Hal ini dibuktikan dengan besarnya jumlah UMKM online berkisar 10 ribu dibanding dengan UMKM fisik hanya berjumlah 4552 UMKM. Namun secara kualitas, pemanfaatan platform online oleh UMKM kategori fashion muslim pada shopee belum cukup baik karena dapat dilihat bukti masih besarnya jumlah UMKM online yang berjenis Regular Seller sebanyak 950 toko dibanding Star Seller hanya 210 toko yang notabene label untuk UMKM yang aktif dan mampu melayani dengan baik. Kemudian didukung juga dari sisi UMKM Lokal yang secara kualitas masih belum baik, merujuk hasil bukti dari rata-rata rating dan follower yang lebih rendah daripada UMKM yang menggunakan platform online.

Kata Kunci:, UMKM, Fesyen Muslim, Web Scrapping, Data Preprocessing, Data Validation, Visualisasi data

Halaman ini sengaja dikosongkan

UTILIZATION ANALYSIS OF ONLINE PLATFORM TO DESCRIBE THE AVAILABILITY OF BUSINESS INFORMATION IN SMEs MUSLIM FASHION IN SURABAYA

Name : Syafrie Dwi Faisal
NRP : 0521 16 4000 0134
Department : Information System FTIK-ITS
Supervisor I : Rully Agus Hendrawan, S.Kom, M.Eng

ABSTRACT

At present many retail MSMEs have taken advantage of the online platform. MSME requires data processing and analysis to keep it updated in knowing the business situation.

The absence of tools to collect data quickly and the lack of knowledge in data processing makes MSMEs using online platforms unable to run optimally. Even though the correct data processing will be information that can be useful for determining decisions and knowing the current state of business in real time. The industry to be chosen is the Muslim fashion industry.

First, data scrapping is done to retrieve data from Tokopedia, Shopee, Google Business and Instagram online platforms. The results of the data are then preprocessed so that the data becomes normal. Data is integrated so that it can be viewed online, offline or online / offline. Followed by the assessment of data quality by comparing the results of integration and selecting the results of the assessment with the highest average precision score. Then a descriptive descriptive analysis is performed by visualizing information.

The results of this study are in the form of a data scheme, a data collection of Muslim fashion MSMEs on several online platforms, a validity test workflow to assess data reliability and information visualization for data analysis. The results of information visualisation prove that in terms of quantity, MSMEs are already good in utilizing online platforms. This is evidenced by the large number of online MSMEs around 10 thousand compared to physical MSMEs which only amount to 4552 MSMEs. But in quality, the use of online platforms by the MSME category of Muslim fashion is not good enough. See evidence of the still large number of online MSMEs that are of Regular Seller type compared to Star Sellers which incidentally are labels for MSMEs that are active and able to serve well.

Keywords: SMEs, Fashion Muslim, Web Scrapping, Data Preprocessing, Data Validation & Data Visualization.

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertandatangan di bawah ini:

Nama : Syafrie Dwi Faisal
NRP : 05211640000134
Tempat/Tanggal lahir : Surabaya/25 Mei 1998
Fakultas/Departemen : Fakultas Teknologi Elektro dan Informatika Cerdas/Sistem Informasi
Nomor : 031-8703152/085730622350/
Telp/Hp/Email : syafriedwif@gmail.com

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

ANALISIS PEMANFAATAN PLATFORM ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA

Bebas Dari Plagiarisme Dan Bukan Hasil Karya Orang Lain.

Apabila dikemudian 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.

10 Agustus 2020



Syafrie Dwi Faisal
NRP.05211640000134

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 ONLINE UNTUK MENGGAMBARKAN KETERSEDIAAN INFORMASI BISNIS PADA UMKM KATEGORI FASHION MUSLIM DI SURABAYA”.

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. Ibu Mahendrawathi ER. S.T., M.Sc., Ph.D. selaku Ketua Departemen Sistem Informasi ITS Surabaya.
3. Rully Agus Hendrawan, S.Kom, M.Eng selaku dosen pembimbing yang telah mencerahkan segenap tenaga, waktu dan pikiran dalam penelitian ini, serta memberikan motivasi yang membangun.
4. Ibu Erma Suryani, ST, MT, 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.
5. Segenap dosen dan karyawan Departemen Sistem Informasi.
6. Orang tua penulis, yang tiada hentinya mendoakan dan memberikan dukungan kepada penulis.
7. Nidia Avisyah Putri, yang banyak berkontribusi dalam melengkapi Buku Tugas Akhir.
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, Mei 2020

Syafrie Dwi Faisal

DAFTAR ISI

LEMBAR PENGESAHAN.....	i
LEMBAR PERSETUJUAN.....	iii
ABSTRAK	v
ABSTRACT	ix
SURAT PERNYATAAN BEBAS PLAGIARISME	11
KATA PENGANTAR	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL.....	xxii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Permasalahan.....	3
1.4 Tujuan	3
1.5 Metode	4
1.6 Manfaat	4
1.7 Target Luaran.....	4
1.8 Relevansi.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Sebelumnya.....	6
2.2 Usaha Mikro Kecil Menengah (UMKM).....	10
2.3 Analitik Bisnis	10
2.3.1 Analitik Deskriptif.....	11
2.4 Data Collection	11
2.4.1 Web Scrapping	11
2.5 Data Pre-processing	12
2.6 Entity Resolution	12
2.7 Data Validation	12
2.8 Data Visualization.....	12
BAB III METODOLOGI	14
3.1 Diagram Metodologi.....	14
3.2 Uraian Metodologi	18
3.2.1 Studi Literatur	18
3.2.2 Data Scrapping	18
3.2.3 Data Preprocessing	21

3.2.4	Data Validation.....	22
3.2.5	Data Loading	23
3.2.6	Pembuatan Metrics Visualisasi.....	23
3.2.7	Visualisasi Data.....	23
3.2.8	Penyusunan Tugas Akhir.....	23
BAB IV ANALISIS PERANCANGAN	25
4.1	Data Scrapping.....	25
4.2	<i>Data Preprocessing</i>	25
4.1	26	
4.2	26	
4.2.1	Data Cleaning.....	26
4.2.2	Data Transformation	26
4.2.3	Data Integration.....	26
4.3	<i>Data Validation</i>	28
4.4	Pembuatan Metrics Visualisasi	31
4.5	Visualisasi Data.....	34
BAB V IMPLEMENTASI	37
5.1	Data Scrapping.....	37
5.1.1	Tokopedia.....	39
5.1.2	Shopee	44
5.1.3	Instagram.....	48
5.1.4	Google Business	49
5.2	<i>Data Preprocessing</i>	51
5.2.1	Tokopedia.....	51
5.2.2	Shopee	54
5.2.3	Google Business	55
5.2.4	Instagram	57
5.2.5	Data Integration.....	57
5.3	Data Validation	60
5.4	Visualisasi Data.....	64
5.4.1	Card	64
5.4.2	Gauge.....	65
5.4.3	Scatter Plot	65
5.4.4	Bar Chart	66
5.4.5	Map.....	66
5.4.6	Donut Chart	67
5.4.7	Wordcloud	68

BAB VI HASIL DAN PEMBAHASAN	68
6.1 Hasil Data.....	68
6.1.1 Hasil Data Marketplace	68
6.1.2 Hasil Data Google Business	80
6.1.3 Hasil Data Instagram.....	84
6.2 Hasil Data Pre-processing.....	85
6.2.1 Hasil Data Pre-Processing Tokopedia.....	85
6.2.2 Hasil Data Pre-Processing Shopee	86
6.2.3 Hasil Data Pre-Processing Google Business ..	87
6.2.4 Hasil Data Pre-Processing Instagram	88
6.2.5 Hasil Data Integration	89
6.3 Data Validation	91
6.4 Hasil Visualisasi Data dan Analisis	91
BAB VI Kesimpulan dan Saran	97
7.1 Kesimpulan	97
7.2 Saran	99
DAFTAR PUSTAKA	102
LAMPIRAN A Kode Program	106
LAMPIRAN B Hasil data.....	112
LAMPIRAN C Data Pendukung	118
LAMPIRAN D Tabel Chi Kuadrat	120
BIODATA PENULIS	121

Halaman ini sengaja dikosongkan.

DAFTAR GAMBAR

Gambar 1.1 Kerangka Kerja Riset Laboratorium Sistem <i>Enterprise</i>	5
Gambar 3. 1 Metodologi Penelitian Bagian 1	14
Gambar 3. 2 Metodologi Penelitian Bagian 2	15
Gambar 3. 3 Metodologi Penelitian Bagian 3	15
Gambar 3. 4 Metodologi Bagian 4	16
Gambar 3. 5 Metodologi Penelitian Bagian 5	17
Gambar 4. 1 Rumus sequencematcher	27
Gambar 4. 2 Rumus Intersection of word	27
Gambar 4. 3 Flowchart data validation	28
Gambar 4. 4 Rumus <i>Krejcie dan Morgan</i>	29
Gambar 4.5 Rumus Precision.....	31
Gambar 5. 1 Visualisasi card.....	65
Gambar 5. 2 Visualisasi Gouge card.....	65
Gambar 5. 3 Visualisasi scatter plot.....	66
Gambar 5. 4 Visualisasi Bar Chart.....	66
Gambar 5. 5 Visualisasi Map	67
Gambar 5. 6 Visualisasi Donut chart	67
Gambar 5. 7 Visualisasi Wordcloud	68
Gambar 6. 1 Hasil Scrapping Data Kategori Tokopedia.....	69
Gambar 6. 2 <i>wordcloud</i> produk Tokopedia.....	70
Gambar 6. 3 Hasil Scrapping Data Produk Tokopedia	70
Gambar 6. 4 Grafik perbandingan jumlah UMKM pengguna promo	71
Gambar 6. 5 Data URL UMKM pada Tokopedia	71
Gambar 6. 6 Grafik trend UMKM setiap tahun mulai dari 2009-2019.....	72
Gambar 6. 7 Data UMKM pada Tokopedia.....	73
Gambar 6. 8 Data Kategori Shopee.....	74
Gambar 6. 9 <i>wordcloud</i> Shopee	75

Gambar 6. 10 Data Produk Shopee	75
Gambar 6. 11 Data URL UMKM Shopee	76
Gambar 6. 12 Grafik jumlah UMKM setiap sub kategori fashion muslim.....	77
Gambar 6. 13 Data URL UMKM Hasil Pre-processing.....	78
Gambar 6. 14 Grafik jumlah UMKM berdasarkan jenis UMKM	79
Gambar 6. 15 Grafik perbandingan jumlah UMKM	79
Gambar 6. 16 Data UMKM Shopee	80
Gambar 6. 17 Data Kecamatan dan Kelurahan	81
Gambar 6. 18 Data Kategori.....	82
Gambar 6. 19 Data Kategori Hasil pre-processing.....	82
Gambar 6. 20 Data UMKM Google Business.....	83
Gambar 6. 21 Hasil data scrapping instagram.....	84
Gambar 6. 22 Hasil data scrapping shopee	85
Gambar 6. 23 Hasil Data Pre-Processing Tokopedia	86
Gambar 6. 24 Ringkasan data UMKM pada data Pre-Processing Shopee	87
Gambar 6. 25 Hasil Data Pre-Processing Google Business ...	88
Gambar 6. 26 Data Pre-Processing Instagram.....	88
Gambar 6. 27 Data Pre-Processing Shopee Instagram.....	89
Gambar 6. 28 Hasil Data Integration Tokopedia, shopee dan instagram	90
Gambar 6. 29 Hasil Data Integration Google My Business,Tokopedia dan Shopee	91
Gambar 6. 30 Laman awal Dashboard	92
Gambar 6. 31 Visualisasi general comparison	93
Gambar 6. 32 Visualisasi halaman UMKM fisik	94
Gambar 6. 33 Visualisasi pada halaman UMKM Tokopedia.	95
Gambar 6. 34 Visualisasi Pada Halaman UMKM Shopee.....	95

Halaman ini sengaja dikosongkan.

DAFTAR TABEL

Tabel 2. 1 Penelitian Sebelumnya	6
Tabel 4. 1 Sample data yang akan diambil.....	30
Tabel 4. 2 Rancangan Matriks Visualisasi	31
Tabel 4. 3 Rancangan luaran visualisasi data.....	35
Tabel 5. 1 Detil Perangkat.....	37
Tabel 5. 2 Tools dasar pembuatan kode scrapping.....	37
Tabel 5. 3 Kueri xpath list kategori Tokopedia	39
Tabel 5.4 Kueri xpath informasi Produk	40
Tabel 5.5 Kueri Xpath Informasi UMKM Tokopedia.....	42
Tabel 5. 6 Kueri Xpath Untuj Mengambil Informasi Kategori Shopee	45
Tabel 5. 7 Kueri Xpath Untuk Mengambil Informasi Produk Shopee	45
Tabel 5. 8 Kueri Xpath Untuk Mengambil URL UMKM Shopee	46
Tabel 5. 9 Kueri Xpath Informasi UMKM Shopee	47
Tabel 5. 10 Kueri Xpath Pada Instagram	49
Tabel 5. 11 Kueri Xpath Informasi UMKM Google My Business.....	50
Tabel 5. 12 Data Preprocessing Tokopedia.....	52
Tabel 5. 13 Data preprocessing shopee	54
Tabel 5. 14 Data Preprocessing Google Business	56
Tabel 5. 15 Data Preprocessing Instagram	57
Tabel 5. 16 Data Integration.....	58
Tabel 5. 17 Threshold Tokopedia Shopee	59
Tabel 5. 18 Threshold Google business, Tokopedia dan shopee	59
Tabel 5. 19 Hasil integrasi data penggunaan beberapa library	59
Tabel 5. 20 penghitungan jumlah data sample	60
Tabel 5. 21 Pengecekan Hasil Data Validasi.....	62
Tabel 5. 22 Hasil Penghitungan Presisi.....	63

Tabel A. 1 Scrapping Script Kategori Fashion Muslim Tokopedia.....	106
Tabel A. 2 Scrapping Script Produk Fashion Muslim Tokopedia	106
Tabel A. 3 Kode preprocessing URL Produk.....	106
Tabel A. 4 Scrapping Script Informasi Kategori Fashion Muslim Tokopedia.....	106
Tabel A. 5 Scrapping Script Kategori Fashion Muslim Shopee	107
Tabel A. 6 Scrapping Script Produk Fashion Muslim Shopee	107
Tabel A. 7 Scrapping Script Data URL UMKM Fashion Muslim Shopee.....	107
Tabel A. 8 Scrapping Script Informasi UMKM Fashion Muslim Shopee.....	107
Tabel A. 9 Scrapping Script Informasi UMKM Google My Business.....	108
Tabel A. 10 Scrapping Script Informasi Instagram.....	108
Tabel A. 11 Kode Preprocessing Informasi UMKM Fashion Muslim Tokopedia	108
Tabel A. 12 Kode Preprocessing Informasi UMKM Fashion Muslim Tokopedia	109
Tabel A. 13 Kode Preprocessing Data Akun Instagram	109
Tabel A. 14 Kode Preprocessing Similarity Data Akun Instagram.....	109
Tabel A. 15 Kode Preprocessing Informasi UMKM Google My Business.....	109
Tabel A. 16 Kode Integrasi Data UMKM Antara Tokopedia Dan Shopee	110
Tabel A. 17 Kode Integrasi Data UMKM Antara Tokopedia, Shopee, dan Google My Business.....	110
Tabel B. 1 Hasil Data Kategori Fashion Muslim Tokopedia	112
Tabel B. 2 Hasil Data Produk Fashion Muslim Tokopedia .	112
Tabel B. 3 Hasil Data URL UMKM Hasil Preprocess Kategori Fashion Muslim Tokopedia.....	112
Tabel B. 4 Hasil Data Informasi UMKM Kategori Fashion Muslim Tokopedia	112
Tabel B. 5 Hasil Data Kategori Fashion Muslim Shopee	113

Tabel B. 6 Hasil data produk fashion muslim shopee	113
Tabel B. 7 Hasil Data URL UMKM Fashion Muslim Shopee	113
Tabel B. 8 Hasil Data Informasi UMKM Fashion Muslim Shopee	113
Tabel B. 9 Hasil Data Instagram Shopee.....	114
Tabel B. 10 Hasil Data Instagram Tokopedia	114
Tabel B. 11 Hasil Data Informasi UMKM Google My Business	114
Tabel B. 12 Data Informasi UMKM Fashion Muslim Tokopedia Hasil Preprocess	114
Tabel B. 13 Data Informasi UMKM Fashion Muslim Shopee Hasil Preprocess	115
Tabel B. 14 Data Informasi UMKM Fashion Muslim Google My Business Hasil Preprocess	115
Tabel B. 15 Data Akun Instagram dan Tokopedia Hasil Preprocess.....	115
Tabel B. 16 Data Akun Instagram dan Shopee Hasil Preprocess	115
Tabel B. 17 Data UMKM Fashion Muslim Hasil Integrasi Antara Tokopedia Dengan Shopee	116
Tabel B. 18 Data UMKM Fashion Muslim Hasil Integrasi Antara Tiga Platform, Google My Business, Tokopedia Dan Shopee	116
Tabel C. 1 Data Kecamatan dan Kelurahan	118
Tabel C. 2 Data Kategori Google Business.....	118
Tabel C. 3 Data Preprocessing Pemilihan Kategori	118
 Tabel D. 1 Tabel Chi Square	120

Halaman ini sengaja dikosongkan

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 peranan penting dalam pertumbuhan ekonomi negara. Dalam perekonomian Indonesia peran UMKM dapat dilihat dari: (1) kedudukan nya sebagai pemain utama dalam kegiatan ekonomi di berbagai sektor, (2) penyedia lapangan kerja yang terbesar, (3) pengembangan kegiatan ekonomi dan local dan pemberdayaan masyarakat, (4) pencipta pasar baru dan sumber inovasi, serta (5) sumbangannya dalam menjaga neraca pembayaran melalui kegiatan ekspor [1]. Angka UMKM dari tahun ke tahun juga cenderung meningkat. Data dari Kementerian Koperasi dan Usaha Kecil Menengah (KUKM) pada tahun 2018 menunjukkan bahwa terdapat sekitar 64 juta pelaku UMKM di Indonesia. Sejauh tahun 2015, UMKM telah memberikan kontribusi pada Produk Domestik Bruto (PBD) sebesar 57-60% dan tingkat penyerapan tenaga kerja sekitar 97% dari seluruh tenaga kerja nasional [2].

Dengan berkembangnya era digital pada sektor ekonomi, sebanyak 9,4 juta UMKM telah beralih dengan menggunakan platform online dalam menjalankan kegiatan ekonomi nya. Penggunaan *platform online* memungkinkan UMKM untuk menjangkau pasar yang lebih luas dan memudahkan pengalaman berbelanja bagi konsumen. Selain itu pemanfaatan platform online juga memungkinkan akses ke UMKM dapat dilakukan secara *multi-platform*. Namun, keberadaan platform online khususnya konsep *multi-platform* membuat persaingan antar UMKM semakin kompetitif. Salah satu penelitian yang dilakukan pada perilaku belanja online di Indonesia menyatakan bahwa kepercayaan, harga, kenyamanan, kemudahan dan

ketersediaan merupakan faktor yang paling utama mempengaruhi konsumen belanja di UMKM online [3]. Hal ini membuktikan semakin banyak nilai manfaat *multi-platform* yang ditawarkan oleh UMKM maka kecenderungan pelanggan memilih UMKM tersebut juga akan meningkat.

Dalam memaksimalkan peluang penggunaan *multi-platform*, UMKM dapat melihat kondisi saat ini. Salah satu cara untuk mengetahui kondisi saat ini adalah dengan melakukan analisis data. Menurut Patton, analisis data adalah proses mengatur data, mengorganisasikan nya kedalam suatu pola, kategori dan satuan uraian dasar [4]. Data yang dianalisis dengan benar dapat menghasilkan informasi untuk membantu proses pengambilan keputusan.

Salah satu sektor industri UMKM yang memiliki potensi untuk memanfaatkan hasil analisis data adalah sektor fesyen terutama fesyen muslim. Tercatat bahwa dari 750 ribu UMKM kategori fesyen yang ada di Indonesia, sekitar 30% merupakan industri fesyen muslim berdasarkan data dari Kemenperin tahun 2016. Berdasarkan data Organisasi Konferensi Islam(OKI), saat ini ekspor fesyen muslim Indonesia berada di peringkat ketiga dengan nilai US\$7,18 Miliar, setelah Bangladesh senilai US\$ 22 miliar dan Turki senilai US\$ 14 miliar [5]. Disamping tingginya nilai ekspor fesyen muslim, Indonesia merupakan negara dengan populasi masyarakat beragama muslim terbesar sehingga memiliki potensi pasar yang besar dan peluang bisnis yang menjanjikan bagi industry fesyen muslim [6].

Untuk mendukung kebutuhan perencanaan analisis data yang baik, dibutuhkan juga ketersediaan dan pengumpulan data yang baik. Analisis data yang akan dilakukan adalah untuk mengetahui kondisi UMKM kategori fesyen muslim di Surabaya dengan menggunakan data yang didapat dengan cara *Web Scrapping*. *Web Scrapping* adalah suatu metode untuk melakukan ekstraksi data dari situs *website*, secara cara cepat, efisien dan otomatis dengan menawarkan data dalam format yang lebih terstruktur [7]. Hasil analisis data dilakukan dengan cara abstraksi menjadi informasi visual.

Hasil dari penelitian ini berupa dokumen hasil analisis data kondisi pemanfaatan *platform online* pada UMKM

kategori fesyen muslim di Surabaya. Diharapkan hasil dari penelitian ini dapat digunakan oleh Pemerintah, UMKM dan masyarakat untuk mengetahui kondisi UMKM fesyen muslim di Surabaya pada *platform online* sebagai fakta pendukung kegiatan penelitian lebih lanjut.

1.2 Rumusan Masalah

Rumusan masalah dari pembuatan tugas akhir ini sebagai berikut:

1. Bagaimana cara menyediakan data UMKM *fashion* muslim di Surabaya dalam platform *online*?
2. Bagaimana hasil pencocokan data hasil integrasi data UMKM *fashion* muslim di Surabaya pada platform *online*?
3. Bagaimana kondisi UMKM *fashion* muslim di Surabaya dalam memanfaatkan platform *online*?

1.3 Batasan Permasalahan

Batasan masalah pada tugas akhir ini sebagai berikut.

1. Data yang dikumpulkan merupakan data hasil *web scrapping* UMKM kategori *fashion* muslim di Surabaya
2. Data didapatkan dari Google My Business, Tokopedia, Shopee dan Instagram.
3. Data yang didapat digunakan untuk keperluan *Descriptive Analytics*.

1.4 Tujuan

Tujuan dari pembuatan tugas akhir ini sebagai berikut:

1. Melakukan pengumpulan dan pengolahan data sehingga tersedia data UMKM *fashion* muslim di Surabaya dari berbagai macam platform *online*.
2. Mengukur hasil pencocokan data platform online dari hasil integrasi data dan informasi sehingga terdapat ukuran sejauh mana kebenaran data dapat dipertanggung jawabkan pada UMKM *fashion* muslim di Surabaya.
3. Mengidentifikasi kondisi UMKM *fashion* muslim di Surabaya secara kualitas dan kuantitas dalam

memanfaatkan platform *online* dalam bentuk visualisasi informasi.

1.5 Metode

Metode yang digunakan pada penelitian ini, yaitu *Web Scrapping*, *Data Preprocessing*, *Data Validation*, *Data Visualization* & *Descriptive Analysis*.

1.6 Manfaat

Berikut manfaat yang diharapkan dari pengerjaan tugas akhir ini:

1. Bagi Peneliti

Sebagai referensi untuk melakukan analisis lebih lanjut mengenai UMKM kategori *fashion* muslim di Surabaya. Alur pengerjaan dan hasil validitas data hasil integrasi dapat digunakan sebagai kerangka untuk melakukan pengukuran kebenaran data sehingga dapat diketahui sejauh mana hasil data yang didapatkan dapat diandalkan untuk melakukan penelitian.

2. Bagi Pemerintah

Sebagai informasi kondisi UMKM kategori *fashion* muslim di Surabaya. Sehingga pihak pemerintah dapat melakukan analisis lebih lanjut baik secara diagnostik, preskriptif dan prediktif. Hasil analisis juga dapat digunakan sebagai fakta pendukung untuk pengambilan keputusan.

1.7 Target Luaran

Target luaran dalam penelitian ini terbagi menjadi beberapa bagian yaitu sebagai berikut:

1. Teknis

Pada bagian teknis, penelitian ini ditargetkan berupa skema data, kumpulan data UMKM fashion muslim pada beberapa platform online dan alur pengerjaan dalam menilai efektifitas data hasil integrasi dengan melakukan uji validitas hingga hasil uji untuk menilai kehandalan data.

2. Visualisasi

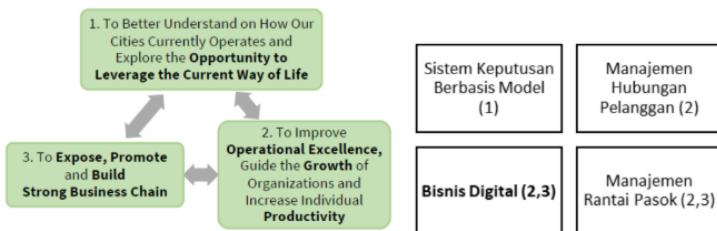
Pada bagian visualisasi, penelitian ini ditargetkan luaran berupa *dashboard* sebagai alat pengelolaan data menjadi informasi visual dan hasil analisis data sebagai metode untuk menjelaskan kondisi data.

3. Dokumen

Pada bagian dokumen, ditargetkan luaran berupa laporan dokumen 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 meningkatkan eksistensi operasional, memandu pertumbuhan organisasi dan produktivitas individu; Untuk mengekspos, mempromosikan, dan membangun rantai bisnis yang kuat. Pada **Gambar 1.1** Kerangka Kerja Riset Laboratorium Sistem Enterprise 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 meliput penelitian-penelitian sebelumnya, kajian pustaka, dan metode yang digunakan selama penggerjaan.

2.1 Penelitian Sebelumnya

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

Tabel 2. 1 Penelitian Sebelumnya

Penelitian 1 [8]	
Judul Penelitian	Implementasi <i>Web Scrapping</i> untuk Pengambilan Data pada Situs <i>Marketplace</i>
Penulis	Dhita Deviacita Ayani, Helen Sasty Pratiwi, Hafiz Muhardi
Tahun	2019
Deskripsi Umum Penelitian	Pada penelitian tersebut menjelaskan tentang pencarian di beberapa situs marketplace mengenai produk yang hendak dibeli secara manual. Para peneliti ingin mencari produk terbaik berdasarkan jumlah produk dengan penjualan terbanyak atau terlaris.

	<p>Dengan bantuan aplikasi <i>web</i> pencarian terbaik dan teknik <i>web scraping</i> ini akan mampu melakukan pencarian di beberapa situs marketplace dan menampilkan hasil pencarian secara bersamaan. Menurut hasil pengujian dengan menggunakan <i>white box testing</i> dan <i>black box testing</i>, sistem telah mampu memberikan hasil terbaik produk dari gabungan hasil pencarian di tiga situs <i>web marketplace</i> sesuai kata kunci yang dimasukkan oleh pengguna.</p> <p>Penelitian ini digunakan sebagai bahan referensi dalam mengetahui secara teknis kegiatan <i>web scrapping</i></p>
Penelitian 2 [9]	
Judul Penelitian	Frameworks for entity matching: A comparison
Penulis	Hanna Kopcke, Erhard Rahm
Tahun	2009
Deskripsi Umum Penelitian	Pada penelitian ini dilakukan analisis komparatif menggunakan 11 kerangka kerja yang berbeda untuk melakukan <i>entity matching</i> atau metode similaritas data. Beberapa parameter

	<p>penilaian dilihat dari hasil jumlah similaritas, efektivitas nilai pengukuran validasi hasil integrasi atau matching dan efisiensi waktu saat melakukan integrasi data.</p> <p>Penelitian ini digunakan sebagai bahan referensi urgensitas pendukung pada hasil tugas akhir ini.</p>
Penelitian 3 [10]	
Judul Penelitian	Real Time Product Analysis using Data Mining
Penulis	Jawahire Nakash, Shaikh Anas, Siddiqi Muzammil Ahmad, Ansari Mohd. Azam, Prof Tabrez Khan
Tahun	2015
Deskripsi Umum Penelitian	Pada penelitian menjelaskan tentang Analisis Produk secara <i>Real-Time</i> menggunakan metode <i>data mining</i> memungkinkan pembeli untuk membandingkan produk dari beberapa situs <i>web E-commerce</i> yang berbeda, sehingga memudahkan mereka untuk membeli produk dengan harga terbaik. Untuk mendapatkan penawaran terbaik dari situs

	<p><i>web E-commerce</i> yang berbeda, digunakan metode <i>web crawler</i> untuk mengambil URL produk.</p> <p>Sebagai referensi untuk mengetahui dan melakukan kegiatan <i>web scrapping</i> secara teknis.</p>
Penelitian 4 [11]	
Judul Penelitian	Web Scraping and Data Acquisition Using Google Scholar
Penulis	D.Pratiba,Abhay M S, Akhil Dua,Giridhar K Shanbhag, Neel Bhandari, Utkarsh Singh
Tahun	2018
Deskripsi Umum Penelitian	<p>Pada artikel tersebut menjelaskan bagaimana melakukan <i>Web Scraping</i> pada data Google Scholar menggunakan python sebagai basis pemograman dan mysql sebagai database. Digunakan Teknik HTML Parsing dengan memanfaatkan URL halaman website. Kemudian data akan disimpan pada format excel.</p> <p>Hasil dan pembahasan dari paper tersebut akan dijadikan referensi oleh penulis untuk memahami langkah-langkah</p>

	dalam melakukan <i>Data Scraping</i> dari website.
--	--

2.2 Usaha Mikro Kecil Menengah (UMKM)

Usaha Mikro, Kecil Menengah (UMKM) adalah pengembangan empat kegiatan ekonomi utama yang menjadi motor penggerak pembangunan Indonesia yaitu Industri Manufaktur, Agribisnis, Bisnis Kelautan dan Sumber Daya Manusia [12]. Menurut Undang-undang Republik Indonesia No.20 tentang UMKM. Kriteria yang digunakan untuk mendefinisikan UMKM seperti yang tercantum dalam pasal 6 adalah nilai kekayaan bersih atau nilai aset tidak termasuk tanah dan bangunan tempat usaha, atau hasil penjualan tahunan. Detil kriteria sebagai berikut:

- a. Usaha mikro adalah unit usaha yang memiliki aset paling banyak Rp.50.000.000, (lima puluh juta rupiah) tidak termasuk tanah dan bangunan tempat usaha dengan hasil penjualan tahunan paling besar Rp.300.000.000 (tiga ratus juta rupiah).
- b. Usaha kecil dengan nilai aset lebih dari Rp. 50.000.000(lima puluh juta rupiah) sampai dengan paling banyak Rp.500.000.000(lima ratus juta rupiah) tidak termasuk tanah dan bangunan tempat usaha memiliki hasil penjualan tahunan lebih dari Rp.300.000.000(tiga ratus juta rupiah) hingga maksimum Rp.2.500.000.000(2 setengah miliar rupiah).
- c. Usaha menengah adalah perusahaan dengan nilai kekayaan bersih lebih dari Rp.500.000.000(lima ratus juta rupiah) hingga paling banyak Rp.10.000.000.000 (sepuluh miliar rupiah) hasil penjualan tahunan di atas Rp.2.500.000.000 (dua setengah miliar rupiah) sampai paling tinggi Rp.50.000.000.000(lima puluh miliar rupiah).

Secara spesifik peneliti akan meneliti UMKM pada kategori fesyen muslim yang ada di Surabaya.

2.3 Analitik Bisnis

Analitik Bisnis adalah suatu studi yang berfokus dengan pengakuan dan penyelesaian masalah berbasis bukti yang terjadi dalam konteks situasi bisnis [13]. Terdapat beberapa pola pikir analitik

pada ilmu analitik bisnis seperti *Descriptive Analytics*, *Prescriptive Analytics* dan *Predictive Analytics*. Pada Tugas Akhir tersebut, peneliti akan menggunakan pola pikir *Descriptive Analytics*. Berikut penjelasan lebih detail terkait pola pikir tersebut

2.3.1 Analitik Deskriptif

Analitik deskriptif menggunakan data historis sebagai masukan untuk menghasilkan statistik dan analisis deskriptif. Dengan kata lain, menceritakan tentang apa yang sedang terjadi. Pada analitik ini, tidak ada pernyataan mengenai probabilitas, hanya ada statistik deskriptif, *Key Performance Indicator* (KPI), dan lain-lain. Analitik ini berkaitan dengan *Business Intelligence* (BI) [14].

2.4 Data Collection

Data Collection adalah proses mengumpulkan dan mengukur informasi tentang variabel yang diminati, dengan cara sistematis yang mapan yang memungkinkan seseorang untuk menjawab pertanyaan penelitian yang dinyatakan, menguji hipotesis, dan mengevaluasi hasil. Tujuan untuk semua pengumpulan data adalah untuk menangkap bukti kualitas yang kemudian diterjemahkan menjadi analisis data yang kaya dan memungkinkan pembangunan jawaban yang meyakinkan dan kredibel untuk pertanyaan yang diajukan [15]. Beberapa metodologi *data collection* seperti wawancara, kuisioner atau dengan menggunakan otomasi teknologi untuk mengambil data di Internet yaitu *Web Scrapping*. Berikut penjelasan terkait metode *Web Scrapping*:

2.4.1 Web Scrapping

Web Scraping adalah Suatu metode untuk melakukan ekstrasi data dari situs *website*, secara cara cepat, efisien dan otomatis dengan menawarkan data dalam format yang lebih terstruktur dan lebih mudah digunakan ,baik untuk proses B2B atau B2C [16]. Penggunaan teknik tersebut biasanya digunakan untuk mengumpulkan data dalam jumlah yang besar. Terdapat beberapa teknik seperti dalam *Web Scrapping* seperti *HTTP Programming*, *DOM Parsing*,*HTML Parsing*, *Web Scraping Software* dan *Computer vision web-page analyser*.

2.5 Data Pre-processing

Data Preprocessing atau *data preparation* adalah sekumpulan proses berulang untuk memanipulasi data mentah, yang sering tidak terstruktur dan berantakan, ke dalam bentuk yang lebih terstruktur dan berguna yang siap untuk analisis lebih lanjut [17]. Pemrosesan data termasuk *data cleaning*, *data reduction*, *data transformation* dan *data integration*. *Data Cleaning*, beberapa aktivitas pada tahap ini seperti menghilangkan nilai-nilai data yang salah, memperbaiki kecacauan data dan memeriksa data yang tidak konsisten. Kemudian yang kedua adalah *Data Integration*, yaitu Menggabungkan data dari beberapa sumber (database, data cube, atau file) ke dalam penyimpanan data yang sesuai. Lalu ada *Data Transformation*, yaitu dengan melakukan normalisasi dan pengumpulan data agar data menjadi sama dan tidak terdapat *outlier*, dan yang terakhir adalah *Data Reduction*, yaitu dengan menguraikan data ke dalam dimensi yang lebih kecil untuk melihat perspektif data yang lebih detail.

2.6 Entity Resolution

Entity Resolution adalah proses mengidentifikasi dan menggabungkan data yang merujuk pada atribut atau entitas yang sama di sumber data yang berbeda [18]. Tahap kegiatan *entity resolution* sangat erat dengan kegiatan *data integration*.

2.7 Data Validation

Data validation adalah sebuah aktivitas yang bertujuan untuk memverifikasi apakah nilai data asli yang diberikan dapat diterima [19]. Kegiatan ini dilakukan untuk menetapkan keabsahan data agar data yang diperoleh tidak *invalid*(cacat). Dalam praktiknya data yang telah dilakukan integrasi akan dilakukan *crosscheck* untuk menilai efektifitas pengukuran. Beberapa *standard* pengukuran seperti menggunakan *precision* dan *recall* [9].

2.8 Data Visualization

Data Visualization adalah suatu proses untuk menafsirkan dan membandingkan data untuk melakukan komunikasi dengan jelas ide-ide kompleks dari data, dengan demikian memudahkan melakukan identifikasi dan analisis pola yang bermakna [20]. *Data Visualization*

menggunakan grafik komputer untuk menunjukkan pola, tren, dan hubungan antar elemen data [21].

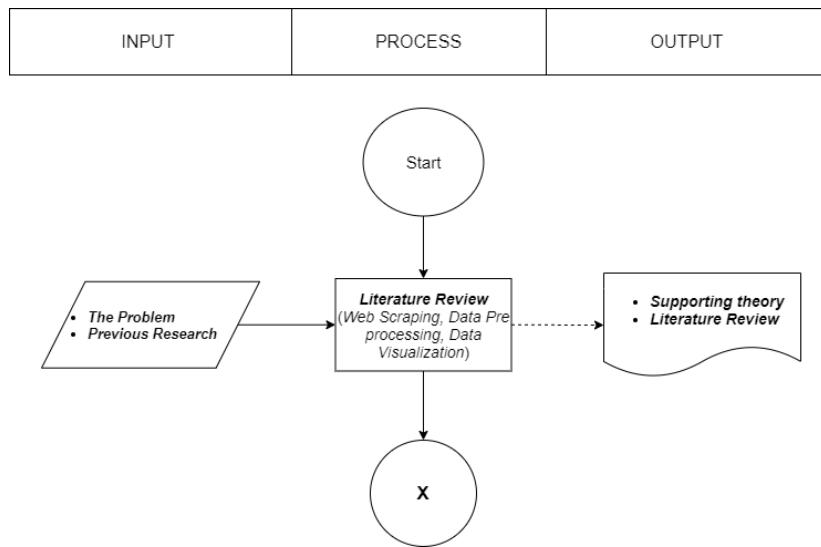
BAB III

METODOLOGI

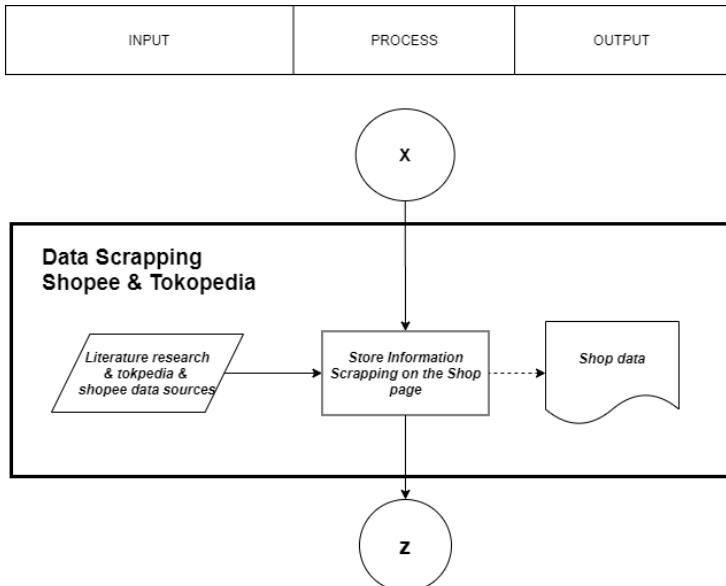
Pada bab ini dijelaskan metodologi yang digunakan sebagai panduan untuk menyelesaikan penelitian tugas akhir ini. Metodologi berdampak pada proses penggerjaan yang terstruktur dan terarah.

3.1 Diagram Metodologi

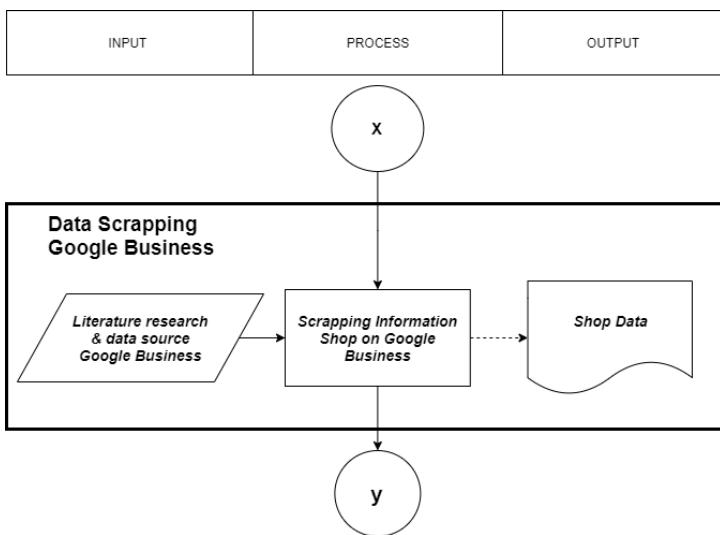
Sub-bab ini menjelaskan tahapan dan proses yang dilakukan dalam penelitian ini. Rangkaian tahapan dan prosesnya terlihat dalam metodologi penelitian yang disusun secara berurutan masing-masing pada Error! Reference source not found., Error! Reference source not found., Error! Reference source not found., Error! Reference source not found. dan **Gambar 3.5** merupakan diagram metodologi penelitian ini.



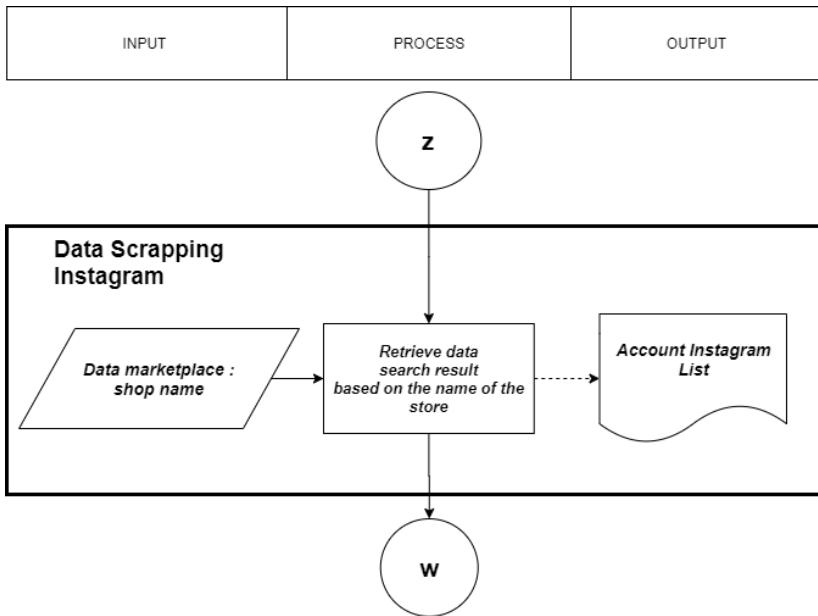
Gambar 3.1 Metodologi Penelitian Bagian 1



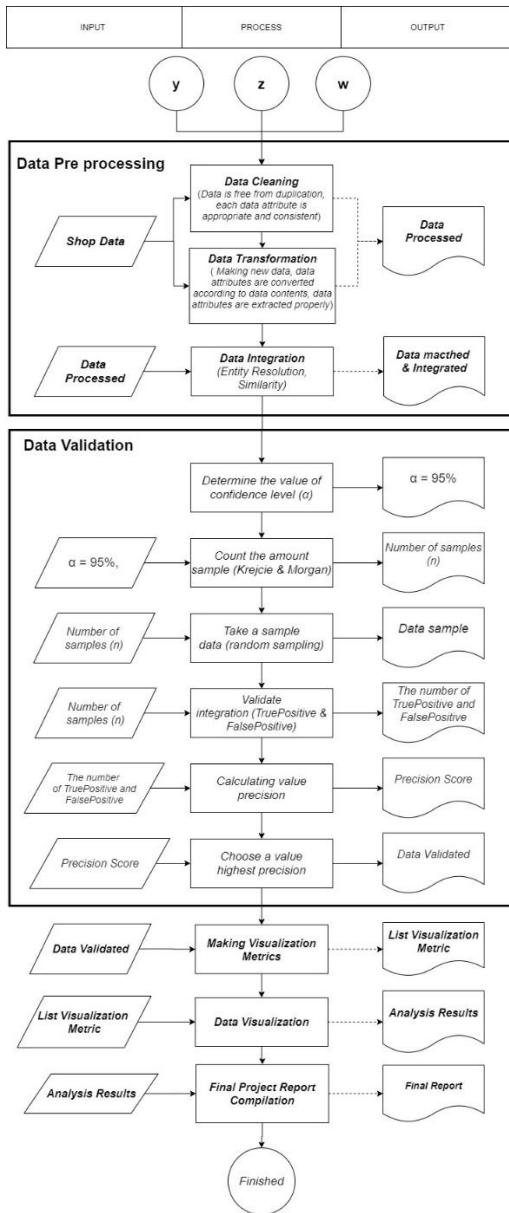
Gambar 3. 2 Metodologi Penelitian Bagian 2



Gambar 3. 3 Metodologi Penelitian Bagian 3



Gambar 3.4 Metodologi Bagian 4



Gambar 3. 5 Metodologi Penelitian Bagian 5

3.2 Uraian Metodologi

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

3.2.1 Studi Literatur

Tahap pertama dalam metodologi Tugas Akhir ini adalah melakukan pencarian studi literatur yang mendukung dan digunakan pada penelitian ini sesuai dengan permasalahan yang diangkat. Identifikasi dan perumusan masalah dilakukan berdasarkan studi literatur yang akan dibahas mengenai *Web Scrapping, Data Preprocessing, Data Integration* dan Visualisasi data.

3.2.2 Data Scrapping

Setelah melakukan pengkajian literatur, proses selanjutnya adalah, proses *data scrapping*. *Data Scrapping* dilakukan pada beberapa platform *platform online* yaitu Tokopedia, Shopee dan Google Business. Adapun proses *data scrapping* pada setiap platform tersebut, yaitu:

1. Scrapping data Marketplace

Kegiatan *data scrapping* 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 scrapping* pada platform *e-commerce* terdiri dari beberapa tahap, yaitu:

a. Scrapping kategori dan sub kategori fesyen muslim

Tahap pertama *data scrapping* pada platform *marketplace* adalah mengambil data kategori beserta sub kategori nya. Kategori yang dipilih adalah kategori fesyen muslim. Sebelum melakukan pengambilan data perlu dilakukan pemahaman struktur website dan lokasi elemen data yang akan diambil. Informasi yang akan diambil adalah nama kategori dan URL kategori. Hasil dari aktivitas ini adalah list data kategori beserta sub kategori dan URL kategori.

b. *Scrapping* Informasi produk

Tahap kedua *data scrapping* pada platform *marketplace* adalah mengambil informasi produk. Pada proses ini dibutuhkan data URL kategori pada proses sebelumnya untuk mengakses setiap produk pada setiap kategori. Informasi yang akan diambil adalah nama produk dan URL produk. Hasil dari aktivitas ini adalah list data produk dan URL produk.

c. *Scrapping* URL UMKM

Tahap ketiga *data scrapping* pada platform *marketplace* adalah mengambil URL UMKM. Pada proses ini dibutuhkan data URL produk untuk bisa mengakses halaman produk. Informasi yang akan diambil adalah URL UMKM pada setiap laman produk. Setelah data didapatkan, URL UMKM perlu dilakukan *preprocessing* untuk menghilangkan duplikasi pada URL UMKM sehingga menghasilkan data yang *unique*. Hasil dari aktivitas ini adalah *list* URL UMKM.

d. *Scrapping* informasi UMKM

Tahap keempat *data scrapping* pada platform *marketplace* adalah mengambil informasi UMKM. Pada proses ini dibutuhkan data URL UMKM untuk dapat mengakses halaman UMKM. Informasi yang diambil pada setiap *platform* berbeda sesuai data yang ditampilkan. Hasil dari aktivitas ini adalah *list* informasi UMKM pada *platform marketplace*.

2. Scrapping data *Google Business*

Kegiatan *data scrapping* untuk mendapatkan informasi UMKM pada *platform google business* membutuhkan data pendukung seperti data kategori google business dan data kecamatan atau kelurahan di Surabaya. Kemudian *scrapping* informasi UMKM dapat dengan baik dilakukan. Adapun proses *data scrapping* pada *platform google business* terdiri dari beberapa tahap, yaitu:

- a. Mengumpulkan data kategori dan data kelurahan/kecamatan
Tahap pertama *data scrapping* pada platform google business adalah mengumpulkan data pendukung kegiatan *data scrapping*. Data tersebut merupakan data kategori google business dan data kecamatan/kelurahan yang didapatkan dari sumber di internet. Adapun hasil dari aktivitas ini adalah list data kategori google business dan data kelurahan/kecamatan yang masih mentah.
 - b. Memilih list kategori yang relevan dengan kategori fesyen muslim
Tahap kedua *data scrapping* pada platform google business adalah memilih list kategori yang relevan dengan kategori fashion muslim. Pemilihan kategori dilakukan secara manual dan *semi-preprocessing*. Kategori yang dianggap tidak relevan akan dibuang dan tersisa kategori yang relevan dengan kategori *fashion* muslim. Hasil dari aktivitas ini adalah *list* data kategori google business relevan dengan kategori fashion muslim.
 - c. *Scraping* informasi UMKM pada Google Business
Tahap ketiga *data scrapping* pada platform google business adalah *scraping* informasi UMKM pada google business. Pada tahap ini diperlukan data input dari data kecamatan/kelurahan dengan data kategori google business yang sudah dipilih. Penggunaan kedua data tersebut digunakan untuk membuat *list keyword* yang akan dieksekusi pada kolom *search* google business. Hasil pencarian setiap UMKM yang muncul akan diambil informasi nya seperti nama, alamat, *link website*, kategori, jam buka, *link shopee*, link Tokopedia dan *link Instagram*. Hasil dari aktivitas ini adalah list informasi UMKM pada google business.
3. *Scraping* data *Instagram*
Kegiatan scrapping di Instagram menggunakan data UMKM *marketplace* yaitu atribut *name* (nama UMKM). Kegiatan ini dengan tujuan mengambil data akun UMKM berdasarkan nama UMKM yang ada pada *marketplace*. Hasil dari aktivitas ini adalah data akun instagram.

3.2.3 Data Preprocessing

Data yang sudah didapatkan sebelumnya adalah data mentah sehingga perlu melewati beberapa proses sampai data siap digunakan. Oleh karena itu, diperlukan proses *data pre-processing* untuk membuat data menjadi bersih dan siap pakai. Adapun proses *data processing* terbagi menjadi beberapa bagian sebagai berikut:

a. *Data Cleaning*

Tahap pertama adalah *data cleaning*, beberapa aktivitas di tahap ini seperti menghilangkan nilai-nilai data yang salah, memperbaiki kecacauan data dan memeriksa data yang tidak konsisten. Hasil dari tahap ini adalah data bebas duplikasi, setiap atribut data sudah sesuai dan konsisten.

b. *Data Transformation*

Tahap kedua adalah *data transformation*, beberapa aktivitas pada tahap ini adalah melakukan normalisasi dan pengumpulan data agar data menjadi sama dan tidak terdapat *outlier*. Hasil dari tahap ini adalah pembuatan data baru, atribut data terkonversi sesuai potensi isi data, dan atribut data terekstrak dengan baik.

c. *Data Integration*

Tahap ketiga adalah *data integration*, yaitu dengan menggunakan metode *entity resolution*. *Entity resolution* dilakukan dengan menggunakan teknik *similarity* atau melihat kemiripan atribut atau entitas pada UMKM dalam data pada platform yang berbeda dengan menggunakan library python *sequencematcher*. Dalam menerapkan teknik *similarity* diperlukan adanya *threshold* atau ambang batas untuk menilai bahwa hasil *similarity* bernilai baik. Nilai *Threshold* didapatkan dengan cara menghitung jumlah benar atau *True Positive (1,00)* dan disesuaikan dengan kondisi data. Nilai *similarity* dari data yang memiliki nilai sama atau lebih dari *threshold* dianggap *match* atau sama.

3.2.4 Data Validation

Tahap *data validation* digunakan untuk memastikan jika data yang diintegrasi memiliki tingkat validitas yang dapat dipertanggung jawabkan. Akan diambil secara data yang didapatkan dengan menggunakan rumus krejcie dan morgan, kemudian dilakukan penghitungan dengan menggunakan rumus *precision*. Berikut detil tahap *data validation*:

1. Menentukan nilai *confidence level*

Nilai *confidence level* digunakan sebagai acuan seberapa yakin atau percaya diri peneliti dengan rentang hasil yang akan dicapai. Semakin tinggi confidence level maka semakin rendah nilai presisi yang akan didapatkan. Confidence level akan digunakan sebagai bantuan dalam menghitung jumlah *sample*.

2. Menghitung jumlah *sample*

Penentuan jumlah *sample* menggunakan rumus krejcie dan morgan. Margin error didapatkan dari penentuan nilai *confidence level* terlebih dahulu. Hasil penghitungan *sample* akan digunakan untuk mengambil data pada hasil integrasi

3. Mengambil *sample* data

Pengambilan *sample* dilakukan sebanyak jumlah hasil penghitungan jumlah *sample*. Pengambilan data *sample* dilakukan secara *random sampling*

4. Melakukan validasi integrasi

Sample data kemudian dilakukan crosscheck secara manual di data lapangan untuk mengetahui apakah hasil integrasi benar-benar sama atau salah. Hasil yang benar akan dinilai sebagai *true positive* dan hasil yang salah akan dinilai sebagai *false positive*

5. Menghitung nilai presisi

Setelah didapatkan hasil data *true positive* dan *false positive*. Dilanjutkan dengan menghitung nilai presisi dengan menghitung hasil yang benar-benar sama dibagi dengan seluruh data hasil integrasi yang memiliki nilai similaritas.

6. Memilih nilai presisi tertinggi

Hasil skor presisi kemudian dibandingkan dan dipilih skor yang memiliki nilai yang paling tinggi. Skor tersebut akan dijadikan acuan sebagai nilai representatif hasil data integrasi.

3.2.5 Data Loading

Setelah data terintegrasi satu sama lain, maka tahap selanjutnya melakukan *data loading* ke tools visualisasi data dan analisis. *Tools* yang akan digunakan adalah PowerBI.

3.2.6 Pembuatan Metrics Visualisasi

Tahap berikutnya adalah Pembuatan metrics. 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 *list* tujuan informasi visual yang akan ditampilkan pada *tools* visualisasi.

3.2.7 Visualisasi Data

Tahap selanjutnya adalah visualisasi data. *Data loaded* pada *tools* kemudian diolah menjadi informasi visual seperti *diagram* atau *chart* sesuai *metrics* visualisasi yang telah dibuat. Kemudian dilakukan analisis secara deskriptif untuk mengetahui kondisi data yang telah didapatkan. Hasil dari analisis adalah kesimpulan tentang kondisi UMKM pada kategori fesyen muslim di Surabaya sesuai data yang didapatkan.

3.2.8 Penyusunan Tugas Akhir

Tahap penyelesaian menghasilkan dua luaran, yaitu laporan tugas akhir dan jurnal mahasiswa. Adapun proses untuk menghasilkan luaran akan dijelaskan sebagai berikut.

b. Menyusun Laporan Tugas Akhir

Penyusunan laporan tugas akhir bertujuan untuk mendokumentasikan langkah-langkah serta hasil analisis dan perhitungan dalam penelitian ini. Selain itu, penyusunan ini bertujuan untuk digunakan sebagai referensi secara praktis

maupun teoritis dalam seleksi. Luaran proses ini berupa dokumen tugas akhir yang dibuat dalam sebuah buku.

c. Menyusun Jurnal Mahasiswa ITS

Penyusunan jurnal mahasiswa ITS bertujuan untuk mendokumentasikan dan mempublikasikan langkah-langkah penggerjaan penelitian ini. Luaran proses ini berupa dokumen jurnal mahasiswa ITS yang diunggah ke dalam portal.

BAB IV

ANALISIS PERANCANGAN

Pada bab ini akan dibahas mengenai seluruh proses analisis dan rancangan untuk *data scrapping*, *data pre processing*, pembuatan metrics visualisasi dan visualisasi data.

4.1 Data Scrapping

Tahap *data scrapping* dilakukan dengan tujuan akhir untuk mengambil informasi UMKM pada platform online. Platform online meliputi shopee, tokopedia, instagram dan google business. Dalam mendapatkan keseluruhan informasi UMKM, proses *data scrapping* perlu melalui beberapa sesuai struktur website dari setiap *platform online*. Dalam penelitian ini proses *data scrapping* setidaknya dibedakan menjadi 2 berdasarkan fungsi nya. Jenis yang pertama berfungsi sebagai penghasil data untuk mencapai data utama. Jenis yang kedua berfungsi sebagai penghasil data informasi UMKM dan digunakan untuk kegiatan analisis secara deskriptif kondisi UMKM kategori *fashion muslim* di Surabaya. Pembuatan dasar program *scrapping* dilakukan oleh **Ubai Yusuf Siraj Samudera** [22] dan disesuaikan berdasarkan kebutuhan peneliti.

4.2 Data Preprocessing

Dari seluruh informasi UMKM yang didapatkan kemudian dilakukan pemrosesan agar data konsisten dan terintegrasi satu sama lain. Secara umum dalam pemrosesan data terdapat 3 tahap, yaitu *data cleaning* dengan contoh menghilangkan duplikasi data dan *handle* data yang kosong, dilanjutkan *data transformation* dengan contoh kegiatan, ekstrasi informasi dari suatu atribut, mengkoversi data sesuai tipe data dan pembuatan data baru dan dilanjutkan dengan *data integration* yaitu seluruh data diintegrasikan satu sama lain agar menjadi satu kesatuan yang siap pakai.

Data Preprocessing data dilakukan pada setiap platform dan sesuai kondisi data yang didapatkan. Pembuatan

kode *data preprocessing* dilakukan oleh **Ubai Yusuf Siraj Samudera** [22]. Sedikit banyak tidak banyak dijelaskan pembuatan kode secara detil.

4.2.1 Data Cleaning

Proses *data cleaning* dilakukan untuk membersihkan data sehingga seluruh atribut data menjadi konsisten. Beberapa perlakuan seperti mengganti nilai NAN menjadi nilai sesuai isi tipe data dan menghilangkan duplikasi data.

4.2.2 Data Transformation

Pada tahap data transformation dilakukan untuk mengkonsolidasikan data ke bentuk yang lebih relevan sesuai isi data. Pada penelitian beberapa rancangan proses *data transformation* antara lain seperti melakukan ekstraksi informasi dari atribut *description* menjadi *COD*, *phone*, *website* dan *instagram* dan melakukan konversi data misal dari atribut *operational_hours* menjadi *operational_hours_duration* untuk melihat data secara numerik.

4.2.3 Data Integration

Seluruh data informasi UMKM dari ketiga platform akan di integrasikan 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. Penggunaan Library *python* yaitu *SequenceMatcher* dan teknik *Intersection of Words* akan digunakan dalam penerapan teknik *similarity*. Kedua teknik akan dipilih salah satu dengan hasil nilai validitas yang paling tinggi. Berikut penjelasan singkat tentang kedua teknik tersebut:

1. SequenceMathcer

Dasar penghitungan *SequenceMatcher* berdasarkan algoritma *Ratcliff-Obershelp similarity algorithm* [23]. Penghitungan dilakukan dengan cara

menghitung dua kali lipat jumlah pencocokan karakter dibagi dengan total jumlah karakter dalam dua string, rumus sesuai **gambar**.

$$d_{ro} = \frac{2c_m}{|s_1| + |s_2|}$$

Gambar 4. 1 Rumus sequencematcher

Dengan penjelasan konotasi sebagai berikut :

C_m = jumlah karakter yang sama

$|s_1| + |s_2|$ = Jumlah karakter kata pertama dan kata target

2. Intersection of Words

Metode *Intersection of Words* adalah metode similaritas dengan menghitung rasio antara jumlah kata yang paling banyak muncul dengan jumlah maksimum dari perbandingan dua string [23]. Rumus dapat dilihat pada **gambar**.

$$i(string_1, string_2) = \frac{\text{common(words}(string_1), \text{words}(string_2))}{\text{max(words}(string_1), \text{words}(string_2))}$$

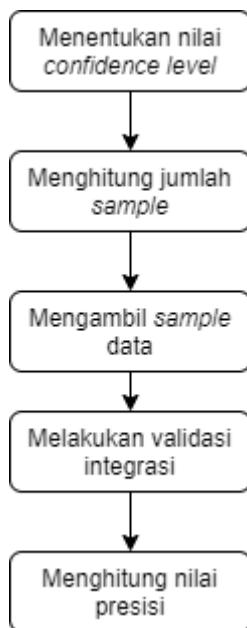
Gambar 4. 2 Rumus Intersection of word

Dibutuhkan juga *threshold* atau ambang batas untuk menilai bahwa hasil *similarity* dinilai baik. Similaritas atribut data akan diterapkan prioritasasi secara urut dari yang tinggi hingga yang paling rendah masing-masing yaitu *link, phone, website, instagram* dan *name*. Setiap atribut memiliki nilai *threshold* yang berbeda disesuaikan dengan kondisi data. Nilai *threshold* dihitung dengan nilai benar atau *TruePositive(1.00)*. Hasil *similarity* yang bernilai sama atau lebih besar dari nilai *threshold* akan dianggap sama atau *match* setiap baris.

4.3 Data Validation

Data validation dilakukan untuk melakukan validasi dari data yang telah dilakukan integrasi. Pada prosesnya kegiatan *data validation* digunakan untuk mempertanggung jawabkan seberapa *valid* hasil integrasi data dengan melakukan survei data secara langsung. Pada penilitian ini dilakukan *data validation* pada dua data hasil integrasi yaitu data tokopedia, shopee dan instagram untuk menilai UMKM secara *online*, data google business, Tokopedia, shopee dan instagram untuk menilai UMKM secara fisik atau *offline* dan *online*.

Terdapat beberapa langkah dalam menghitung *data validation*. Detil langkah-langkah ada pada **Gambar 4.1**.



Gambar 4.3 Flowchart data validation

a. Menentukan nilai *confidence level*

Confidence level diperlukan untuk mengukur sejauh apa data dapat dipercaya kebenerannya. Pada penelitian ini ditentukan *confidence level* sebesar 95%.

b. Menghitung jumlah *sample*

Dalam melakukan perhitungan *sample* digunakan rumus *Krejcie dan Morgan* seperti **Gambar 4.4**.

$$n = \frac{\chi^2 \cdot N \cdot P(1 - P)}{(N - 1) \cdot d^2 + \chi^2 \cdot P(1 - P)}$$

Gambar 4. 4 Rumus *Krejcie dan Morgan* [24]

Dengan penjelasan konotasi sebagai berikut:

χ^2 = Nilai Chi Kuadrat dk = 1 (Lihat pada **LAMPIRAN D**)

P = Proporsi Populasi

d = Tingkat kesalahan 5% (0,05)

N = Nilai populasi (Hasil dari total yang match)

Nilai N adalah hasil total data integrasi yang dianggap *match* atau yang memiliki nilai similaritas.

c. Mengambil *sample* data

Pengambilan *sample* dilakukan secara acak pada dua data hasil integrasi yaitu Tokopedia, shopee dan google business, shopee dan tokopedia. *Sample* data hasil integrasi pertama diambil dengan melihat UMKM Tokopedia yang memiliki akun shopee, kemudian pada hasil integrasi kedua melihat data pada google business yang memiliki akun tokopedia, google business yang memiliki akun shopee dan

google business yang memiliki akun Tokopedia dan shopee. Detil *sample* data yang akan diambil bisa dilihat pada **Tabel 4.1**:

Tabel 4. 1 Sample data yang akan diambil

Sumber data	Data sample (Memiliki atribut link)
Google Business	Tokopedia
Google Business	Shopee
Google Business	Tokopedia, shopee
Tokopedia	Shopee

Proses pengambilan data disesuaikan dengan jumlah penghitungan masing-masing jumlah *sample*. Data yang diambil adalah data yang *match* atau sama yang memiliki nilai similaritas.

d. Melakukan validasi integrasi

Proses validasi integrasi dilakukan untuk memastikan *sample* data hasil integrasi benar-benar memiliki data yang cocok sesuai hasil similaritas. Validasi integrasi dilakukan secara manual dengan secara langsung melakukan survei pada sumber data di Internet sesuai jumlah *sample* data yang telah ditentukan.

Dari hasil validasi akan diketahui berapa banyak data yang benar-benar memiliki akun *marketplace*. Jika dalam proses validasi tidak ditemukan akun *link* sesuai hasil similaritas atau akun UMKM malah tidak benar, maka UMKM tersebut dianggap tidak valid atau bernilai *False Positive*.

e. Menghitung nilai *precision*

Hasil validasi integrasi kemudian dilakukan penghitungan nilai presisi. Presisi digunakan untuk menilai ketepatan antara informasi dengan jawaban yang diberikan oleh hasil integrasi. Detil rumus presisi sesuai pada **Gambar 4.3**.

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

Gambar 4.5 Rumus Precision [25]

Dengan penjelasan konotasi sebagai berikut :

TP (True Positive) = Hasil data benar cocok pada survei manual.

TP (True Positive) + FP (False Positive) = Hasil total match atau yang mengandung nilai similaritas dan atribut *link* sesuai data integrasi.

4.4 Pembuatan Metrics Visualisasi

Pada tahap ini dilakukan pemetaan *metrics* visualisasi berdasarkan analitik dari data. Pembuatan metrics digunakan sebagai panduan dalam pembuatan visualisasi. Terdapat tiga data yang didapatkan dari hasil *preprocessing* termasuk integrasi yaitu data integrasi dengan menggabungkan data google business, shopee dan Tokopedia, data *preprocess* shopee dan Tokopedia. Dilakukan pengamsusian data dengan membagi rencana metrics menjadi empat bagian platform yaitu UMKM online, UMKM secara fisik *offline* dan *online*. Penjabaran detil definisi platform dapat dilihat pada Tabel. Rancangan metrics secara umum dilakukan oleh **Aldy Syah Daviq Ramadhan** [26], dengan terdapat kustomisasi metriks. Hasil rancangan luaran metrics visualisasi yang akan dihasilkan seperti pada **Tabel 4.2**

Tabel 4. 2 Rancangan Matriks Visualisasi

Platform	Atribut	Metrics	Tujuan
----------	---------	---------	--------

UMKM Fisik (Offline)	Index	Jumlah UM KM	Melihat jumlah U MKM pada perspektif tertentu
	Review	Rata-rata review	Melihat bagaiman a nilai tengah dari jumlah review pada perspektif tertentu
	Operational_ hours	Rata- rata jumlah ja m kerja	Melihat intensitas aktivitas kerja UM KM pada perspektif tertentu
	Rating Rating,Oper ational_our s	Rata-rata rating	Melihat bagaiman a nilai tengah dari rating pada perspektif tertentu
		Korelasi rati ng dengan Jamk erja	Melihat bagaiman a hubungan antara variabel rating dengan jam kerja pada perspektif ter tentu
Online	Index	Jumlah UM KM	Melihat jumlah U MKM pada perspektif ter tentu
	Active_pro duct/ numOf_pro duct	Rata-rata jumlah katal og produk	Melihat nilai tega h dari jumlah produ kyang aktif dijual padaUMKM 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
	Voucher	Jumlah voucher	Mengetahui frekuensi penggunaan voucher pada UMKM
	numOf_Media	Rata-rata media	Melihat sedikit banyak jumlah media yang digunakan pada UMKM
UMKM Fisik (<i>Offline</i>) dan Online	Index	Jumlah UMKM	Melihat jumlah UMKM pada perspektif tertentu
	Review	Rata-rata review	Melihat bagaimana nilai tengah dari jumlah review pada perspektif tertentu

4.5 Visualisasi Data

Setelah Metrics dibuat, selanjutnya ditentukan informasi visual dengan pemilihan penggunaan *chart*, *diagram* ataupun suatu metode visual sesuai dengan *metrics* visualisasi. Penentuan *Key Performance Indicator* dari setiap rencana visualisasi juga dipetakan untuk mengetahui atribut data apa

yang akan digunakan ketika visualisasi data dilakukan. Berikut rancangan luaran visualisasi data yang akan dihasilkan seperti pada **Tabel 4.3**

Tabel 4.3 Rancangan luaran visualisasi data

Platform	Atribut	Metrics	Jenis Visualisasi
Offline	Index	Jumlah UMKM	<i>Card, bar chart, map</i>
	Review	Rata-rata review	<i>Card</i>
	Operational_hours	Rata-rata jumlah jam kerja	<i>Card</i>
	Rating	Rata-rata rating	<i>Gauge</i>
	Rating, Operational_hours	Korelasi rating dengan Jamkerja	<i>Scatter chart</i>
Online	Index	Jumlah UMKM	<i>Card, map</i>
	Active_product/numOf_product	Rata-rata jumlah katalog produk	<i>card</i>
	Follower	Rata-rata jumlah follower	<i>card</i>
	Review	Rata-rata review	<i>card</i>
	Rating	Rata-rata rating	<i>Gauge</i>
	Seller_type	Jumlah UMKM berdasarkan jenis UMKM	<i>Donut chart</i>

	Rating,product_sold	Korelasi rating dengan jumlah produk yang dijual	<i>Bar Chart</i>
	Delivery_services	Mode dari layanan pengiriman	<i>Wordcloud</i>
	Product_category	Mode dari kategori produk	<i>Wordcloud</i>
	Voucher	Jumlah UMKM yang menggunakan voucher	<i>Bar Chart</i>
	numOf_Media	Rata-rata jumlah media yang digunakan	<i>Card</i>
UMKM Fisik (<i>Offline</i>) dan Online	Index	Jumlah UMKM	<i>Card</i>
	Review	Rata-rata review	<i>Pie Chart</i>

BAB V

IMPLEMENTASI

Pada bab ini dijelaskan implementasi kode program dari rancangan yang telah dibuat. Beberapa kode program antara lain untuk pengumpulan, pembersihan, dan visualisasi data.

5.1 Data Scrapping

Data scrapping atau pengumpulan data dilakukan sebagai proses awal untuk mendapatkan data dari platform online. *Data Scrapping* diomotasi menggunakan kode program. Secara keseluruhan, program dijalankan pada dua perangkat yang berbeda. Detil perangkat dan dasar *tools* pendukung untuk pembuatan kode program masing-masing dapat dilihat pada Error! Reference source not found. & Error! Reference source not found..

Tabel 5. 1 Detil Perangkat

Perangkat	Spesifikasi
<i>Local (OS) Operating Systems Windows</i>	Processor Intel i7 7700, RAM 16GB
<i>Virtual Machine (VM) Operating Systems (OS) Linux</i>	Standard D4s v3 (4 vcpus, 16 GiB memory)

Tabel 5. 2 Tools dasar pembuatan kode scrapping

Tools	Nama Tools	Deskripsi
Bahasa Pemrograman	<i>Python</i>	<i>Python programming language</i>
<i>Library</i>	<i>Pandas</i>	Alat mengakses file data sesuai format
	<i>Selenium</i>	<i>Browser automation</i> untuk menjalankan

		<i>driver</i> dan mensimulasikan interaksi elemen
	<i>pyvirtualdisplay</i>	<i>Display simulator</i> pada VM linux agar program berjalan secara <i>headless</i>
	<i>time</i>	<i>Time Running Control</i> untuk monitor waktu Ketika scrapping data dilakukan
Mesin pencari	<i>chromedriver</i>	Standalone server yang mengimplementasikan standard W3C webdriver
Metode <i>Scraping</i>	<i>Headless</i>	VM OS Linux
	<i>Not Headless</i>	Local OS windows
Bahasa Kueri	XPATH	Alat menemukan lokasi elemen data

Luaran tahap *data scrapping* adalah informasi UMKM pada platform online. Platform online meliputi Tokopedia, shopee dan google business. Kode program *scrapping* dibuat berdasarkan data yang akan diambil dan sesuai struktur laman di setiap platform online. Secara umum seluruh kode program diawali dengan *import library*, *loading file* atau memuat data, dilanjutkan dengan kueri lokasi elemen data dengan XPATH dan *saved data* atau menyimpan data yang didapatkan. Adapun proses tersebut dibagi berdasarkan *platform online*, yaitu sebagai berikut:

5.1.1 Tokopedia

Tahap *data scrapping* Tokopedia meliputi data kategori, informasi produk, URL UMKM dan informasi UMKM fashion muslim tokopedia.

1. Scrapping URL kategori dan sub kategori fashion muslim

Dari hasil rancangan scrapping **bagian**, terdapat dua interaksi elemen bertahap yang harus dilakukan sebelum data kategori beserta URL nya didapatkan. Dalam mendapatkan elemen interaksi digunakan fungsi **visibility_of_element_located** karena tipe elemen bersifat *hover*. Berikut merupakan potongan kode interaksi elemen untuk memunculkan *list* kategori seperti pada **Gambar 5.1**.

```

27 #Interaksi elemen pertama
28 wait_kategori_button = WebDriverWait(driver, 3)
29 .until(EC.visibility_of_element_located((By.XPATH, '//*[@class="css-vk082c"]'))))
30
31 ActionChains(driver).move_to_element(wait_kategori_button).perform()
32
33 #Interaksi elemen kedua
34 WebDriverWait(driver, 3)
35 .until(EC.visibility_of_element_located((By.XPATH, '//*[@class="css-me46ht"]'))))

```

Gambar 5. 1 Interaksi elemen URL Kategori Tokopedia

Setelah interaksi elemen dilakukan dan list kategori telah muncul, Langkah berikutnya dilakukan pengambilan informasi kategori sesuai elemen-elemen yang ada. Pengambilan data dilakukan menggunakan kueri xpath sesuai lokasi elemen pada **Tabel 5.3**. Dari hasil yang didapatkan, terdapat **61** nama kategori dan URL kategori fashion muslim Tokopedia. Dengan struktur umum *scraping* yang dijelaskan, maka keseluruhan program *scrapping* informasi kategori Tokopedia dapat dilihat pada **LAMPIRAN A**

Tabel 5. 3 Kueri xpath list kategori Tokopedia

Nama Kolom	Xpath	Tipe Elemen
------------	-------	-------------

kategori	<code>//*[@class="css-me46ht"] //*[@class="css-1qaqbbz"]</code>	<i>hover</i>
url_kategori	<code>//*[@class="css-me46ht"] //*[@class="css-1qaqbbz"]</code>	<i>href</i>

2. Scrapping informasi produk pada setiap kategori dan sub kategori

Scraping informasi produk membutuhkan masukan URL kategori dari data kategori. Pertama produk dilakukan filter lokasi dengan memilih Surabaya sebagai kota. Pemilihan filter lokasi mengakibatkan perubahan pada URL kategori. Dapat diketahui filter lokasi Surabaya adalah **252**. Data informasi produk yang diambil adalah judul produk dan URL produk.

Jumlah katalog produk pada setiap hasil halaman memiliki informasi yang terbatas, sehingga dibutuhkan interaksi elemen untuk pindah ke halaman katalog produk selanjutnya. Hasil Implementasi dan potongan kode masing-masing untuk pengambilan informasi produk dan interaksi elemen untuk pindah ke hamalan berikutnya dengan xpath dapat dilihat pada **Tabel 5.4** dan **Gambar 5.2**. Hasil *scrapping* informasi produk sebanyak **193.804**. Secara keseluruhan kode program dapat dilihat pada **LAMPIRAN A**.

Tabel 5.4 Kueri xpath informasi Produk

Nama Kolom	Xpath	Tipe Elemen
title	<code>//*[@class="css-bk6tzz e1nlzfl3"]//*[@class="css-1bjwylw"]</code>	<i>text</i>
url_produk	<code>//*[@class="css-bk6tzz e1nlzfl3"]/a</code>	<i>href</i>

```

next_page =
driver.find_elements_by_xpath('//*[@i[@class="css-98hn3t e19tp72t1"]]')
if(next_page):
    #
    #Lokasi XPATH elemen data informasi produk
    #
    next_page[0].click()
else:
    break

```

Gambar 5. 2 Interaksi elemen untuk menggeser halaman

3. URL UMKM fashion muslim

Dalam mendapatkan URL UMKM akan digunakan URL produk dengan dilakukan *Preprocessing* data untuk melakukan normalisasi URL. Hasil akhir normalisasi merubah struktur URL dari <https://tokopedia.com/url-UMKM/url-produk>, menjadi <https://tokopedia.com/url-UMKM>. Langkah selanjutnya dilakukan penghapusan duplikasi URL UMKM dengan penggunaan fungsi *drop_duplicates()* hingga URL UMKM bersifat *unique*. Potongan kode *splitting* data dan preprocessing URL UMKM bisa dilihat pada Gambar 5.3. Jumlah Total URL UMKM yang bersifat unik sebanyak **11.176**. Keseluruhan kode program dapat dilihat pada **LAMPIRAN A**

```

15 #get toko link
16 df_url_toko_notpromo["link"] =
17     df_url_produk_notpromo["link"].str.rsplit("/", 1).str[-2]
18
38 #get unique all toko link
39 df_url_toko_all_unique = df_url_toko_all.drop_duplicates('link')
40

```

Gambar 5. 3 Kode Splitting data dan preprocessing URL Tokopedia

4. Scrapping informasi UMKM

Informasi UMKM didapatkan dari setiap URL UMKM. Digunakan fungsi *for* untuk mengakses seluruh data URL UMKM. Selain halaman utama, informasi UMKM pada Tokopedia akan dilakukan beberapa interaksi elemen untuk dapat mengakses informasi UMKM yang lebih lengkap.

Setidaknya ada tiga interaksi elemen yang dilakukan, interaksi pertama dilakukan *pointing* untuk melihat *reputation point*. Interaksi kedua dilakukan *click* pada button info UMKM untuk akses beberapa informasi seperti deskripsi, alamat, waktu UMKM bergabung dan jasa pengiriman. Interaksi terakhir dilakukan *click* untuk melihat statistik UMKM yang didalam nya terdapat informasi seperti rating, statisktik kepuasan, produk aktif, etalase UMKM dan transaksi sukses. Potongan kode dan implementasi masing-masing untuk interaksi elemen dan pengambilan data informasi UMKM menggunakan xpath dapat dilihat pada **Gambar 5.4** dan **Tabel 5.4**. Keseluruhan kode program dapat dilihat pada **LAMPIRAN A**

```
#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()

#open info toko
    info_buttons = driver.find_elements_by_xpath('//*[@[@class="css-rhfifq-unf-btn e1grgruw00"]]')
    driver.execute_script("arguments[0].click();", info_buttons[0])

#open statistics
    statistic_element = driver.find_elements_by_xpath('//*[@[@class="css-16zlw0p-unf-heading e1qvo2ff6"]]')
    driver.execute_script("arguments[0].click();", statistic_element[0])
```

Gambar 5.4 Interaksi elemen pada halaman informasi UMKM

Tabel 5.5 Kueri Xpath Informasi UMKM Tokopedia

Elemen	Xpath	Tipe Elemen
Name	//*[@class="css-1yw0bav"]/span[1]	<i>text</i>
Link	<i>Didapatkan dari data URL UMKM</i>	<i>href</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	<code>//*[@class="css-1ac807k e1hrfe840"]</code>	<i>text</i>
Product_sold	<code>//*[@class="css-lzwncz-unf-heading e1qvo2ff2"]</code>	<i>Text</i>
Rating	<code>//*[@class="css-rfs3ih-unf-heading e1qvo2ff2"]</code>	<i>text</i>
Review	<code>//*[@class="css-1s96mum-unf-heading e1qvo2ff6"]</code>	<i>text</i>
Owner	<code>//*[@class="css-1ag3tdd-unf-heading e1qvo2ff4"]</code>	<i>text</i>
Description	<code>//*[@class="css-f4j8r2-unf-heading e1qvo2ff8"]</code> [0]	<i>text</i>
Address	<code>//*[@class="css-f4j8r2-unf-heading e1qvo2ff8"]</code> [1]	<i>text</i>
Since	<code>//*[@class="css-1c0g9ad e1ufc1ph0"]</code> /p[3]	<i>text</i>
numOf_delivery_services	<code>//*[@class="css-2ibdj3"]</code>	<i>Count/len</i>
Delivery_services	<code>//*[@class="css-15n97o5 e1ufc1ph0"]</code>	<i>text</i>
numOf_satisfied	<code>//*[@class="css-1wem450-unf-heading e1qvo2ff2"]</code>	<i>text</i>
numOf_neutral	<code>//*[@class="css-1jcf8a5-unf-heading e1qvo2ff2"]</code> [0]	<i>text</i>

numOf_unsatisfied	//*[@class="css-1jcf8a5-unf-heading e1qvo2ff2"] [1]	<i>text</i>
Process_speed	//*[@class="css-cy9l5o-unf-heading e1qvo2ff4"]	
numOf_storefront	//*[@class="css-z2vy7e"]	<i>Count/len</i>
Storefront	//*[@class="css-17mrx6g"]	<i>text</i>
Active_product	//*[@class="css-zst5sr-unf-heading e1qvo2ff4"] [2]	<i>text</i>
Success_transaction	//*[@class="css-zst5sr-unf-heading e1qvo2ff4"] [3]	<i>text</i>

5.1.2 Shopee

Beberapa informasi yang akan dikumpulkan pada shopee seperti data kategori, URL produk, URL UMKM hingga informasi UMKM.

1. Scrapping URL kategori fashion muslim

Untuk mendapatkan URL shopee kategori fashion muslim diperlukan akses URL <https://shopee.co.id/Fashion-Muslim-cat.43>. Atribut data yang diambil meliputi nama kategori beserta sub kategori dan URL kategori. Terdapat interaksi elemen dengan fungsi *click* untuk melihat tampilan keseluruhan kategori dan sub kategori. Potongan kode interaksi dapat dilihat pada **Gambar 5.5**.

```
driver.get('https://shopee.co.id/Fashion-Muslim-cat.43')
driver.implicitly_wait(2)
arrow_down_button = driver.find_elements_by_xpath('//*[@class="stardust-icon stardust-icon-arrow-down"]')
arrow_down_button[0].click()
time.sleep(2)
```

Gambar 5.5 Interaksi elemen pada kategori fashion muslim shopee

Hasil pengambilan didapatkan sebanyak **11** kategori *fashion muslim* yang tersedia pada shopee. Implementasi pengambilan informasi kategori dan sub kategori menggunakan xpath bisa dilihat pada **Tabel 5.6**. Kode program secara lengkap dapat dilihat pada **LAMPIRAN A**.

Tabel 5. 6 Kueri Xpath Untuk Mengambil Informasi Kategori Shopee

Nama Kolom	Xpath	Tipe Elemen
kategori	<code>//*[@class="shopee-category-list_main-category_link"]</code> <code>//*[@class="shopee-category-list_sub-category"]</code>	<i>text</i>
url_kategori	<code>//*[@class="shopee-category-list_main-category_link"]</code> <code>//*[@class="shopee-category-list_sub-category"]</code>	<i>href</i>

2. Scrapping informasi produk pada setiap kategori

Dalam mendapatkan informasi produk diperlukan URL kategori. Pengambilan data dilakukan pada setiap URL kategori dan halaman yang tersedia. Pertama produk dilakukan filter lokasi, shopee tidak menyediakan pilihan filter lokasi dengan skala kota oleh karena itu **Jawa Timur** sebagai lokasi filter. Pemilihan filter lokasi mengakibatkan perubahan pada URL kategori. Informasi yang diambil informasi judul produk dan URL UMKM. Implementasi *scrapping* data untuk mengambil informasi produk pada xpath dapat dilihat pada **Tabel 5.7**.

Tabel 5. 7 Kueri Xpath Untuk Mengambil Informasi Produk Shopee

Nama Kolom	Xpath	Tipe Elemen
------------	-------	-------------

title	//*[@class="O6wiAW"]	<i>text</i>
url_produk	//*[@class="col-xs-2-4 shopee-search-item-result__item"]/div/a	<i>href</i>

Katalog produk shopee disusun sedemikin rupa sehingga informasi katalog produk pada setiap halaman terbatas. Dibutuhkan interaksi elemen untuk menggeser ke halaman berikut nya. Potongan kode untuk menggeser ke halaman berikut nya dapat dilihat pada **Gambar 5.6**.

```

while True:
    driver.implicitly_wait(7)
    next_page = driver.find_elements_by_xpath(
        '//button[@class="shopee-button-outline shopee-mini-page-controller__next-btn"]')
    #
    # Kode xpath pengambilan informasi produk
    #
    next_page[0].click()
    else:
        break
index+=1

```

Gambar 5.6 Kode Untuk Menggeser Ke Halaman Berikutnya

Hasil *scrapping* didapatkan sebanyak **7619** produk. Secara lengkap, kode program dapat dilihat pada *scrapping LAMPIRAN A*.

3. Scrapping URL UMKM

Pengambilan URL UMKM dilakukan dengan mengakses setiap URL dari produk. Atribut data yang diambil meliputi adalah URL UMKM yang terdapat pada halaman produk. Implementasi kode xpath untuk mengambil URL UMKM dan kode program secara lengkap masing-masing bisa dilihat pada **Tabel 5.8** dan **LAMPIRAN A**

Tabel 5.8 Kueri Xpath Untuk Mengambil URL UMKM Shopee

Nama Kolom	Xpath	Tipe Elemen
kategori	Didapatkan dari data kategori	-
url_produk	//a[@class=" _136nGn"]	<i>href</i>

4. *Pre-processing URL UMKM*

Data URL UMKM yang berhasil didapatkan kemudian dilakukan *preprocessing data* untuk menghilangkan duplikasi URL dengan menggunakan fungsi *drop_duplicates()* agar URL UMKM bersifat unik satu sama lain. Potongan kode preprocessing dapat dilihat pada **Gambar 5.7**.

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

Gambar 5.7 Kode Preprocessing URL UMKM

Terdapat **3470** UMKM yang *unique*. Secara keseleruhan kode program bisa dilihat pada **LAMPIRAN A**

5. *Scraping informasi UMKM*

Informasi UMKM didapatkan melalui URL UMKM yang didapatkan. Implementasi kode program dilakukan dengan menggunakan Bahasa kueri xpath sesuai pada **Tabel 5.9**. Kode program secara lengkap dapat dilihat di **LAMPIRAN A**.

Tabel 5.9 Kueri Xpath Informasi UMKM Shopee

Elemen	Xpath	Tipe Elemen
link	Didapatkan dari data URL 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	class="section-seller-overview__item-text" and	<i>text</i>

	contains(., "produk")]/div[2]]	
chat_performanc e	class="section-seller- overview__item-text" and contains(., "Performa chat")	<i>text</i>
followers	class="section-seller- overview__item-text" and contains(., "pengikut")	<i>text</i>
rating_review	class="section-seller- overview__item-text" and contains(., "penilaian")	<i>text</i>
since	class="section-seller- overview__item-text" and contains(., "bergabung")	<i>text</i>
voucher	class="shop-page__section shop-page__vouchers"	<i>boolean</i>
numOf_media	class="image- carousel__dots"	<i>len/coun t</i>
description	contains(@class, "shop- page-shop-description")	<i>text</i>

5.1.3 Instagram

Informasi Instagram yang akan diambil meliputi *account*, *description* dan *link*. *Scraping* pada instagram menggunakan informasi nama UMKM pada Tokopedia dan shopee yang akan dilakukan pada kolom pencarian. Terdapat beberapa interaksi elemen yang dibutuhkan sebelum dapat melakukan *scraping* informasi. Beberapa diantara nya seperti : masuk akses akun, *click* elemen untuk mengeluarkan pop up dan *click* kolom *search*. Potongan kode interaksi elemen dapat dilihat pada **Gambar 5.8**

```

#LOGIN
username_element = driver.find_element_by_name('username')
username_element.send_keys("scraper023")

password_element = driver.find_element_by_name('password')
password_element.send_keys("scraper")

password_element.send_keys(Keys.RETURN)
driver.implicitly_wait(15)

#Close Pop up Element
notnow_element = driver.find_element_by_xpath('//*[@[@class="a0OlW_HoLwm "]]')
notnow_element.click()

for each in nama_toko:
    ignored_exceptions=(NoSuchElementException,StaleElementReferenceException)
    WebDriverWait(driver,15,ignored_exceptions
                  =ignored_exceptions).until(expected_conditions.presence_of_element_located
                                              ((By.XPATH, '//*[@class="LWmhU_0aCwM"]')))

#SEARCHs
search_element = driver.find_element_by_xpath('//*[@[@class="LWmhU_0aCwM"]/input')
search_element.clear()
driver.implicitly_wait(15)
search_element.send_keys(str(each))
driver.implicitly_wait(15)
result_element = driver.find_elements_by_xpath('//a[@class="yCE8d "]')

```

Gambar 5. 8 Interaksi Elemen Pada Instagram

Implementasi lokasi elemen pada xpath bisa dilihat pada **Tabel 5.10**. Pada scrapping menggunakan data tokopedia didapatkan sebanyak **214.344** dan shopee didapatkan sebanyak **41.514**. Kode program secara keseluruhan dapat dilihat di **LAMPIRAN A**.

Tabel 5. 10 Kueri Xpath Pada Instagram

Nama Kolom	Xpath	Tipe Elemen
Account	//*[@[@class="Ap253"]]	Text
description	//*[@[@class="Fy4o8"]]	Href
link	//*[@[@class="Fy4o8"]]	Text

5.1.4 Google Business

Informasi UMKM yang akan diambil seperti nama, kategori, jam buka, website, rating, review dan instagram.

1. Menyiapkan data pendukung *scrapping*

Dalam melakukan *data scrapping* di google business, dibutuhkan data pendukung. Data pendukung meliputi data kecamatan/kelurahan dan data kategori google business.

Kedua data tersebut didapatkan dari sumber internet. Kemudian dikonversi menjadi format csv. Data kategori terdapat **475** kategori dan **153** untuk data kecamatan Detil data dapat dilihat pada **LAMPIRAN C**.

2. Memilih kategori relevan dengan fashion muslim

Data kategori google business dilakukan *semi-preprocessing* dengan memilih kategori yang relevan dengan fashion muslim. Beberapa diantara nya seperti *restaurant*, *medical*, *electronic* untuk kategori yang tidak dipilih dan *hijab*, *mukenah*, *outewear* adalah beberapa kategori yang dipilih. hasil *preprocess* list kategori yang lengkap dapat dilihat pada **LAMPIRAN C**.

3. Scrapping informasi UMKM

Dalam penerapan untuk mendapatkan informasi UMKM google business, data kategori yang relevan dan data kecamatan digabungkan kemudian dibuat sebuah list *keyword*. Seluruh *keyword* akan di eksekusi untuk kemudian dilakukan pencarian pada kolom *search*. Implementasi pencarian informasi UMKM menggunakan xpath dan kode program secara lengkap masing-masing dapat dilihat pada **Tabel 5.11** dan **LAMPIRAN C**.

Tabel 5. 11 Kueri Xpath Informasi UMKM Google My Business

Elemen	Xpath	Tipe Eleme n
Name	class="css-1qaqbbz" class="css-1nykm5o"	hover
Website	class="SPZz6b"	text
Rating	class="Ob2kfd"/span/a/spa n	text
numOf_google_re view	class="hqzQac"	text
Price	@class="YhemCb"	text

Category	<i>class="YhemCb"</i>	<i>text</i>
Address	<i>class="LrzXr"</i>	<i>text</i>
Operational_hours	<i>/div/div/div[1]/div[2]/table/tbody</i>	<i>text</i>
Phone	<i>@class="LrzXr zdqRlf kno-f"</i>	<i>text</i>
instagram	<i>class="PZPZlf kno-vrt-t"</i>	<i>text</i>

5.2 Data Preprocessing

Dari seluruh informasi UMKM yang didapatkan kemudian dilakukan pemrosesan agar data konsisten dan terintegrasi satu sama lain. Secara umum dalam pemrosesan data terdapat 3 tahap, yaitu *data cleaning* dengan contoh menghilangkan duplikasi data dan *handle* data yang kosong, dilanjutkan *data transformation* dengan contoh kegiatan, ekstrasi informasi dari suatu atribut, mengkoversi data sesuai tipe data dan pembuatan data baru dan dilanjutkan dengan *data integration* yaitu seluruh data diintegrasikan satu sama lain agar menjadi satu kesatuan yang siap pakai.

Preprocessing data dilakukan pada setiap platform dan sesuai kondisi data yang didapatkan. Pembuatan kode *data preprocessing* dilakukan oleh teman saya bernama **Ubai Yusuf Siraj Samudera** [22]. Sedikit banyak tidak banyak dijelaskan pembuatan kode secara detil.

5.2.1 Tokopedia

Pada *data preprocessing* Tokopedia proses *data cleaning* dan *data transformation* dilakukan pada seluruh atribut informasi UMKM tokopedia. Pada *data cleaning*, beberapa perlakukan seperti mengganti nilai kosong atau *NAN* dengan format ‘No (nama atribut)’ dan menghapus duplikasi pada atribut data. Kemudian pada *data transformation*, data dilakukan *information extraction* pada atribut *description*. Kemudian dihasilkan atribut baru yaitu *COD*, *phone*, *website* dan *instagram*.

Pada *Atribute normalization*, beberapa data yang diubah adalah *followers* dengan merubah tipe data *string* menjadi *int*, merubah isi atribut *phone* agar sesuai dengan nomor universal. Lalu *atribute conversion*, seperti merubah atribut *address*, agar dapat dideteksi secara kecamatan dan kelurahan. Secara keseluruhan detil atribut yang di *preprocess* dan kode program masing-masing dapat dilihat pada **Tabel 5.12** dan **LAMPIRAN A**

Tabel 5. 12 Data Preprocessing Tokopedia

Nama Atribut	Preprocess
Name	Tidak ada <i>preprocess</i>
Link	- Menghapus duplikasi data
Seller_type	- Mengganti <i>nan</i> dengan “Regular seller”
City	- Filter kota di Surabaya
Followers	- Mengkonvensi nilai ‘k’ dan ‘m’ menjadi <i>int</i>
Reputation_point	- Normalisasi & membersihkan data
Product_sold	- Mengkonvensi nilai ‘k’ dan ‘m’ menjadi <i>int</i> - Mengganti nilai kosong jadi ‘0’
Rating	-
Review	- Normalisasi & membersihkan data
Owner	- Mengganti <i>nan</i> dengan “no owner information”
Description	- Ekstrasi informasi menjadi atribut baru yaitu COD, phone, website dan instagram
Address	- Normalisasi & membersihkan data

	<ul style="list-style-type: none"> - Konversi atribut menjadi kecamatan dan kelurahan
Since	<ul style="list-style-type: none"> - Memisah/<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	<ul style="list-style-type: none"> - Merubah tipe data menjadi <i>int</i>
numOf_neutral	<ul style="list-style-type: none"> - Merubah tipe data menjadi <i>int</i>
numOf_unsatisfied	<ul style="list-style-type: none"> - Merubah 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 data yang kosong - Merubah tipe data menjadi <i>int</i>
Success_transaction	<ul style="list-style-type: none"> - Merubah tipe data menjadi <i>int</i>
COD (hasil ekstraksi)	<ul style="list-style-type: none"> - Mengikutkan nilai index
Phone (hasil ekstraksi)	<ul style="list-style-type: none"> - mengkoversi isi atribut <i>phone</i> agar sesuai dengan nomor universal
Website (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Instagram (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Kecamatan (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Kelurahan (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Monthly_joined	Tidak ada <i>preprocess</i>

year_joined	Tidak ada <i>preprocess</i>
-------------	-----------------------------

5.2.2 Shopee

Pada pemrosesan data pada shopee. Secara keseluruhan hampir sama seperti *preprocess* data pada Tokopedia. Kegiatan *Information Extraction* pada atribut *description* menjadi atribut baru yaitu *COD*, *phone*, *website* dan *instagram*. Lalu dilakukan *Atribute normalization* dan *Atribute conversion* pada beberapa atribut. Secara keseluruhan detil atribut yang di *preprocess* dan kode program masing-masing dapat dilihat pada **Tabel 5.13** dan **LAMPIRAN A**

Tabel 5. 13 Data preprocessing shopee

Nama Atribut	Preprocess
Link	- Menjadikan sebagai index
Name	
City	- Filter kota di Surabaya
seller_type	- Mengganti <i>nan</i> dengan “Regular seller”
product_category	- Menghilangkan string ‘Semua Produk\n’
numOf_product	- Mengkonvensi nilai ‘RB’ menjadi <i>int</i> - Merubah tipe data menjadi <i>int</i>
chat_performance	- Memisah/ <i>split</i> data menjadi atribut <i>chat_performance_unit</i> - Konversi data persentase menjadi <i>int</i>
Followers	- Mengkonvensi 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	- Ekstrasi informasi menjadi yaitu <i>COD</i> , <i>phone</i> , <i>website</i> dan <i>instagram</i>
COD (hasil ekstraksi)	- Mengikutkan nilai index
Phone (hasil ekstraksi)	- mengkoversi isi atribut <i>phone</i> agar sesuai dengan nomor universal
Website (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Instagram (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Rating (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Review (hasil ekstraksi)	Tidak ada <i>preprocess</i>
Chat_performance_unit (hasil ekstraksi)	Tidak ada <i>preprocess</i>

5.2.3 Google Business

Pada informasi UMKM google business 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 perlakukan seperti, menambah atribut *operational_hours_duration* dari konversi data dari *operational_hours*.

Ekstrasi informasi dari atribut *website* yang mengandung atribut *e-commerce* menjadi atribut tokopedia, shopee dan instagram. Kemudian melakukan ekstrasi dan konversi informasi dari atribut *address* menjadi atribut kecamatan dan kelurahan. Secara keseluruhan atribut yang dilakukan *preprocess* dan detil kode dapat dilihat pada **Tabel 5.14** dan **LAMPIRAN A**

Tabel 5. 14 Data Preprocessing Google Business

Nama Atribut	Preprocess
Name	Menghapus data yang duplikat
Website	Ekstrasi informasi yang menjadi atribut baru yaitu <i>Tokopedia</i> , <i>Shopee</i> , dan <i>instagram</i>
Rating	Tidak ada <i>preprocess</i>
numOf_google_review	Normalisasi & membersihkan data
Price	Mengganti data yang kosong menjadi 'No Price'
Category	Menghapus kategori yang tidak sesuai
Address	<ul style="list-style-type: none"> - Normalisasi & membersihkan data - Konversi atribut menjadi kecamatan dan kelurahan
Operational_hours	Konversi atribut menjadi <i>operational_hours_duration</i>
Phone	Normalisasi & membersihkan data
Instagram	Menambahkan data hasil ekstraksi informasi
Tokopedia	Normalisasi data
Shopee	Normalisasi data

Operational_hours_duration (Hasil Ekstraksi)	Tidak ada <i>preprocess</i>
Kecamatan (Hasil Ekstraksi)	Tidak ada <i>preprocess</i>
kelurahan (Hasil Ekstraksi)	Tidak ada <i>preprocess</i>

5.2.4 Instagram

Hasil *scrapping* informasi instagram dilakukan *preprocessing*. Pada proses *data cleaning*, perlakuan seperti membersihkan data dari *string* yang mengandung nilai '#', *explore*, dan *location* dilakukan untuk mendapatkan hanya akun instagram. Pada *data transformation* dilakukan similaritas antara nama UMKM *marketplace* dengan akun instagram yang didapatkan menggunakan fungsi *sequencematcher*. Hasil similaritas tertinggi akan dipilih sebagai kandidat akun instagram dengan penentuan *threshold*. *Threshold* digunakan sebagai ambang batas asumsi nilai similaritas dianggap baik. Nilai *threshold* dapat dilihat pada **Tabel 5.15**. Secara keseluruhan kode dapat dilihat pada

Tabel 5. 15 Data Preprocessing Instagram

Platform	Threshold
Instagram – Tokopedia	0.85
Instagram - Shopee	0.85

5.2.5 Data Integration

Terdapat 2 luaran data hasil integrasi dengan masing-masing tujuan untuk menjelaskan kondisi UMKM dari platform *online* dan *offline*. Integrasi data pertama menggunakan data Tokopedia, shopee dan hasil similaritas akun *marketplace* pada instagram, secara umum digunakan untuk melihat kondisi UMKM secara *online*. Integrasi berikutnya digunakan untuk

melihat kondisi UMKM secara *offline* maupun *offline* dan *online*, detil integrasi dapat dilihat pada **Tabel 5.16**.

Tabel 5. 16 Data Integration

Platform	Sumber Data	Data yang akan diintegrasikan
Online	Tokopedia	Instagram
	Shopee	Instagram
	Tokopedia	Shopee
UMKM Fisik (Offline)	Google Business	Tokopedia
UMKM Fisik (Offline) + Online		Shopee

Integrasi dilakukan dengan menggunakan metode *entity resolution* dengan teknik *similarity*. Atribut data yang dipilih antara lain *phone*, *website*, *instagram* dan *name* karena masing-masing *platform* memiliki data tersebut. Prioritasasi juga berlaku pada setiap atribut. Similaritas atribut akan berlaku jika terdapat atribut data yang kosong atau *nan*. Teknik similaritas berlaku dengan adanya *threshold*. Setiap atribut memiliki *threshold*, detil pada masing-masing *threshold* pada kedua data dapat dilihat pada **Tabel 5.17** dan **Tabel 5.18**. Seperti contoh jika similaritas atribut *phone* dijalankan dan melewati *threshold* maka data akan dianggap sama, sebaliknya jika *phone* tidak melewati batas *threshold*, kemudian data yang lain masih kosong maka akan dilakukan similaritas pada atribut berikutnya.

Dilakukan beberapa percobaan pada library yang berbeda yaitu *sequencematcher*, *IntersectionofWord* dan *Max3grams* untuk membandingkan total similaritas yang didapatkan. Jumlah similaritas masing-masing *library* dapat dilihat pada **tabel**. Kode keseluruhan dapat dilihat pada **LAMPIRAN A**

Tabel 5. 17 Threshold Tokopedia Shopee

Prioritasasi	Nama Atribut	Threshold
1	Phone	0.96 / 96%
2	Website	0.86 / 86%
3	instagram	0.83 / 83%
4	name	0.50 / 50%

Tabel 5. 18 Threshold Google business, Tokopedia dan shopee

Prioritasasi	Nama Atribut	Threshold tokopedia	Threshold shopee
1	Phone	0.96 / 96%	0.96 / 96%
2	Website	0.86 / 86%	0.86 / 86%
3	instagram	0.83 / 83%	0.83 / 83%
4	name	0.90 / 90%	0.90 / 90%

Tabel 5. 19 Hasil integrasi data penggunaan beberapa library

Library	Integrasi data	Jumlah Data	Jumlah Similaritas	Execution Time (menit)
Sequence Matcher	Tokopedia-shopee	10731	1162	4,34
	Google Business-Tokopedia Shopee	7934	499	63,39

Intersection of Word	Tokopedia-shopee	10731	71	7,12
	Google Business-Tokopedia Shopee -	7934	30	1,25

5.3 Data Validation

Pada hasil data validation dilakukan sesuai pada **bagian**. Berikut detil hasil *data validation* yang dilakukan:

a. **Confidence Level**

Penentuan *confidence level* ditentukan sebesar 0.9/90%. *Confidence level* digunakan sebagai derajat kepercayaan, seberapa yakin dengan hasil integrasi bahwa data benar-benar sama.

b. **Menghitung jumlah sample**

Penentuan jumlah sample dilakukan sesuai hasil data integrasi dan *library*. Pada **Tabel 5.19** dijelaskan detil penghitungan masing-masing jumlah data *sample*.

Tabel 5. 20 penghitungan jumlah data sample

Library	Data Integrasi	Jumlah Similaritas(N)	$n = \frac{X^2 \cdot N \cdot P(1-P)}{(N-1) \cdot d^2 + X^2 \cdot P(1-P)}$
Sequence Matcher	Google Business - Tokopedia	476	163
	Google Business - Shopee	44	37

	Google Business s - Tokope dia, shopee	21	19
	Tokope dia - Shopee	1162	204
Intersection of Word	Google Business s - Tokope dia	27	10
	Google Business s - Shopee	3	2
	Google Business s - Tokope dia, shopee	0	0
	Tokope dia - Shopee	71	12

c. Mengambil *sample* data

Dari hasil *sample* data (n) dilakukan *random sampling* dengan mengambil data sebesar jumlah yang telah dihitung.

d. Melakukan validasi integrasi

Setiap *sample* data kemudian dilakukan validasi apakah benar masing-masing data hasil integrasi benar-benar

sama sesuai dengan aslinya. Data yang benar-benar sama akan dinilai sebagai data *true positive* dan data yang salah akan dinilai sebagai data *False Positive*. Hasil detil pengecekan data validasi bisa dilihat pada **Tabel 5.20**

Tabel 5. 21 Pengecekan Hasil Data Validasi

Library	Data Integrasi	n	True Positive	False Positive
Sequence Matcher	Google Business – Tokopedia	163	98	65
	Google Business – Shopee	37	25	12
	Google Business - Tokopedia, shopee	19	10	9
	Tokopedia - Shopee	204	134	70
Intersection of Word	Google Business – Tokopedia	10	3	7
	Google Business – Shopee	2	0	2
	Google Business - Tokopedia, shopee	0	0	0

	Tokopedia – Shopee	12	5	7
--	--------------------	----	---	---

e. **Menghitung nilai presisi**

Hasil validasi berupa jumlah nilai *True Positive* dan *False Positive* kemudian dihitung nilai presisi masing-masing data integrasi. Hasil lengkap penghitungan data presisi bisa dilihat pada **Tabel 5.21.**

Tabel 5. 22 Hasil Penghitungan Presisi

Data Integrasi	Memiliki hasil similaritas	True Positif (TP)	False Positif (FP)	Precision ($\frac{TP}{TP+FP}$)
Sequence Matcher	Google Business - Tokopedia	98	65	0,60
	Google Business - Shopee	25	12	0,67
	Google Business - Tokopedia, shopee	10	9	0,52
	Tokopedia - Shopee	134	70	0,66
Average Precision (Sequence Matcher)				0,61 / 61%
Intersection of Word	Google Business - Tokopedia	3	7	0,30

	Google Business - Shopee	0	2	0
	Google Business - Tokopedi a, shopee	0	0	0
	Tokopedi a - Shopee	5	7	0,41
Average Precision (Intersection of Words)				0,1775 / 17%

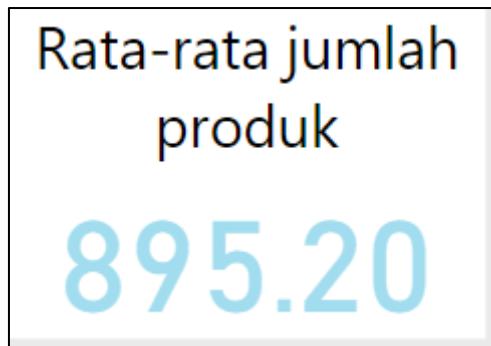
Dari data yang didapatkan dapat ditarik kesimpulan jika hasil integrasi yang dilakukan menggunakan dua metode yang berbeda masing-masing memiliki tingkat validitas rata-rata hasil penghitungan *precision* sebesar **61%** untuk *SequenceMatcher* dan **17%** untuk *IntersectionofWords*.

5.4 Visualisasi Data

Pembentukan visualisasi sesuai pada rancangan visualisasi yang tertera pada **Tabel 4.3**. Berikut merupakan ringkasan implementasi untuk jenis grafik yang dibangun :

5.4.1 Card

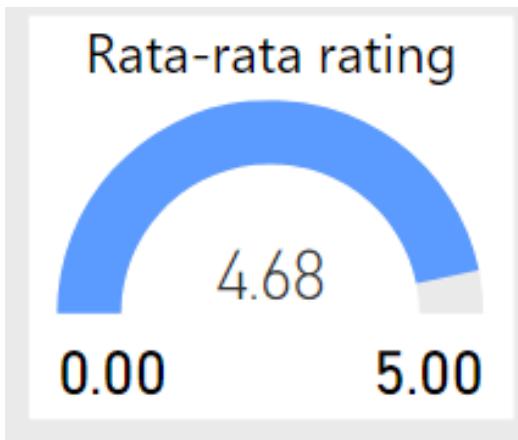
Visualisasi *card* digunakan untuk menunjukkan suatu data bersifat numerik secara ringkas dan jelas. Salah satu implementasi *diagram* sesuai dengan matriks adalah menghitung rata-rata jumlah produk yang dijual. Detil grafik *card* bisa dilihat pada **Gambar 5.9**.



Gambar 5. 9 Visualisasi card

5.4.2 Gauge

Salah satu *chart* yang digunakan pada penelitian ini adalah *Gauge Chart*. *Gauge Chart* berguna untuk membandingkan antara nilai yang memiliki angka kecil. Salah satu implementasi pada data yang didapatkan bisa dilihat pada Gambar 5.10



Gambar 5. 10 Visualisasi Gouge card

5.4.3 Scatter Plot

Implementasi *scatter plot* dalam penelitian yaitu digunakan digunakan adalah pada rata-rata produk yang terjual

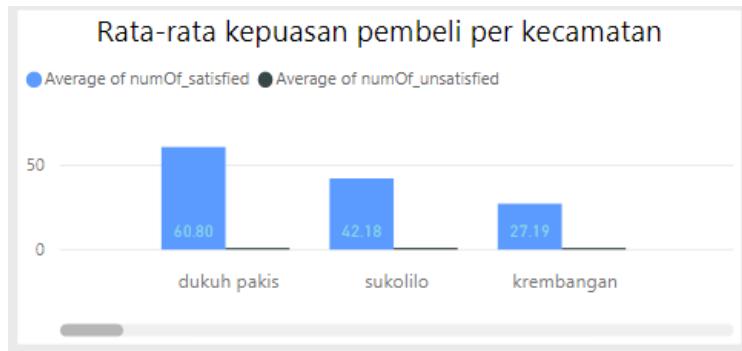
aatau *product_sold* dan *rating*. Visualisasi ini dibangun sesuai yang ditunjukkan pada **Gambar 5.11**.



Gambar 5. 11 Visualisasi scatter plot

5.4.4 Bar Chart

Bar Chart adalah grafik untuk menampilkan data sederhana. Salah satu contoh penerapan *bar chart* pada penelitian ini adalah untuk menampilkan informasi rata-rata kepuasan atau *numOf_satisfied* pada setiap kecamatan dengan bantuan PowerBI, visualisasi ini dapat dibangun dengan langkah yang ditunjukkan pada **Gambar 5.12**

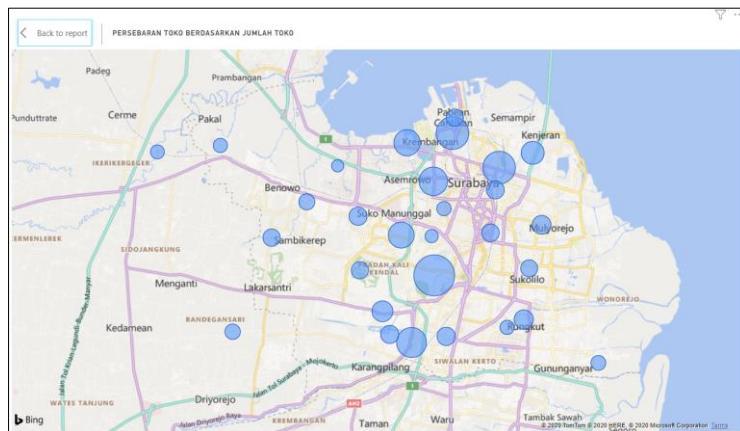


Gambar 5. 12 Visualisasi Bar Chart

5.4.5 Map

Dalam penelitian ini, salah satu implementasi map digunakan untuk menampilkan informasi persentase UMKM berdasarkan jumlah UMKM Dengan bantuan PowerBI,

visualisasi ini dapat dibangun dengan langkah yang ditunjukkan pada **Gambar 5.13**.



Gambar 5. 13 Visualisasi Map

5.4.6 Donut Chart

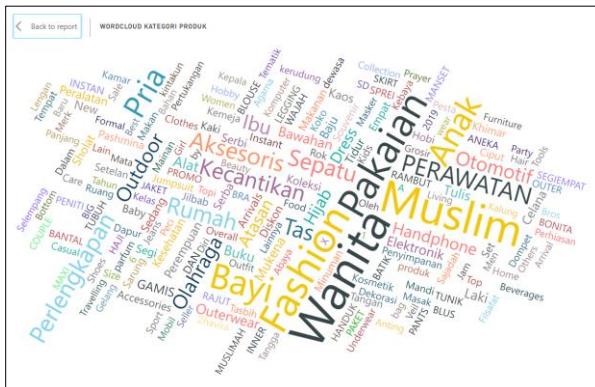
Donut chart digunakan untuk mengerti irisan antara data dalam bentuk persentase namupun numerik. Salah satu implementasi *donut chart* adalah menghitung persebaran jenis UMKM. Bentuk visualisasi ini dapat dilihat pada **Gambar 5.14**.



Gambar 5. 14 Visualisasi Donut chart

5.4.7 Wordcloud

Wordcloud berguna untuk representasi visual dari data teks. Dalam penelitian ini, salah satu implementasi *wordcloud* digunakan untuk menampilkan informasi kategori produk yang dijual pada shopee. Informasi grafik bisa dilihat pada **Gambar 5.15.**



Gambar 5. 15 Visualisasi Wordcloud

BAB VI

HASIL DAN PEMBAHASAN

Bab ini membahas hasil implementasi yang telah dilakukan serta analisis dari proses *data scrapping*, *data processing*, dan *data loading* pada beberapa platform online yang menjadi sasaran studi kasus penelitian ini, yaitu Tokopedia, Shopee dan Google Business.

6.1 Hasil Data

Dalam tahap ini akan dijelaskan uraian terkait hasil data yang didapatkan dari proses *scrapping* pada *marketplace* dan *scrapping* pada Google Business.

6.1.1 Hasil Data Marketplace

Pembahasan hasil proses data *scrapping* pada *marketplace* yang terdiri dari *scrapping* Tokopedia dan Shopee. Pada masing-masing *marketplace* proses *scrapping* dilakukan

untuk menghasilkan beberapa data diantaranya yaitu, kategori dan sub kategori *fashion muslim*, informasi produk, URL UMKM, informasi UMKM yang didapatkan seperti data kategori, informasi produk, URL UMKM dan informasi produk pada Tokopedia.

1. Hasil Scrapping Tokopedia

Beberapa hasil scrapping data Tokopedia seperti data kategori, informasi produk, URL UMKM dan data UMKM

a. Data Kategori Tokopedia

Data kategori yang berhasil didapatkan berjumlah **61** data kategori beserta sub kategori. Keseluruhan data dapat dilihat pada **LAMPIRAN B**. Ringkasan data kategori dapat dilihat pada **Gambar 6.1**

	kategori	link
0	aksesoris muslim	https://www.tokopedia.com/p/fashion-muslim/aks...
1	atasan muslim wanita	https://www.tokopedia.com/p/fashion-muslim/ata...
2	baju renang muslim	https://www.tokopedia.com/p/fashion-muslim/baj...
3	bawahan muslim wanita	https://www.tokopedia.com/p/fashion-muslim/baw...
4	dress muslim wanita	https://www.tokopedia.com/p/fashion-muslim/dre...
5	jilbab	https://www.tokopedia.com/p/fashion-muslim/jilbab
...
56	baju ihram	https://www.tokopedia.com/p/fashion-muslim/per...
57	mukena	https://www.tokopedia.com/p/fashion-muslim/per...
58	rompi sholat	https://www.tokopedia.com/p/fashion-muslim/per...
59	sajadah	https://www.tokopedia.com/p/fashion-muslim/per...
60	sarung	https://www.tokopedia.com/p/fashion-muslim/per...
61	tasbih	https://www.tokopedia.com/p/fashion-muslim/per...

62 rows × 2 columns

Gambar 6. 1 Hasil Scrapping Data Kategori Tokopedia

b. Data Produk Tokopedia

Data produk didapatkan dari input URL kategori. Seluruh produk pada setiap kategori memiliki data sebanyak **193.804** produk. Kemudian dilakukan pembuatan *wordcloud* untuk mengetahui produk yang paling banyak banyak dijual pada di Tokopedia. Sesuai **Gambar 6.2** beberapa kata yang paling besar ada **baju koko, baju muslim , peci dan kopiah**.



Gambar 6. 2 wordcloud produk Tokopedia

Keseluruhan data bisa dilihat pada **LAMPIRAN B**.
Ringkasan data bisa dilihat pada **Gambar 6.3**

kategori	title	link
0 aksesoris muslim	Kaos Kaki Jempol Muslimah Polos Bukan kaos kak...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
1 aksesoris muslim	KAOS KAKI JEMPOL TELAPAK HITAM BAHAN POLYESTER	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
2 aksesoris muslim	PAKET CRAFT BY ALRAYYAN	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
3 aksesoris muslim	Bros cantik dan anggun. Bisa dipakai utk hijab...	https://ta.tokopedia.com/promo/v1/clicks/8a-xg...
4 aksesoris muslim	Masker Perlengkapan Haji Umrah dan Umroh Pri...	https://www.tokopedia.com/tokosurabayaa/masker...
5 aksesoris muslim	Tuspin Mutara	https://www.tokopedia.com/fanilasurabaya/tuspi...
...
193799 tasbih	Tasbih Premium Tiga Daun 33 Butir 8mm Oleh Ole...	https://www.tokopedia.com/toserah/tasbih-premi...
193800 tasbih	tasbih kokka kaukah 99 butir 10.5 mm hitam	https://www.tokopedia.com/emyyokka/tasbih-kok...
193801 tasbih	Tasbih Jari Digital/Penghitung/Tasbih Mini/Fin...	https://www.tokopedia.com/oleh-oleh-hajji/tasbi...
193802 tasbih	Tasbih Batu Cina 99 Per Lusin / 12 Pcs	https://www.tokopedia.com/oleh-oleh-hajji/tasbi...
193803 tasbih	Tasbih Kayu Gaharu 99 Butir Kecil Souvenir Ole...	https://www.tokopedia.com/oleh-oleh-hajji/tasbi...
193804 tasbih	Tasbih Batu Oval 33 Oleh Oleh Haji Umrah Grosi...	https://www.tokopedia.com/toserah/tasbih-batu...

193805 rows x 3 columns

Gambar 6. 3 Hasil Scrapping Data Produk Tokopedia

c. Data URL UMKM pada Tokopedia

Data URL UMKM Tokopedia didapatkan melalui *Splitting* URL produk. URL produk memiliki 2 jenis yaitu URL normal dan URL promo. *Preprocessing* dilakukan terlebih dahulu untuk melakukan normalisasi URL promo menjadi URL produk. Setelah dilakukan *splitting data*, data dilakukan *preprocess* untuk menghilangkan duplikasi UMKM. Terdapat sebanyak **11176** URL UMKM pada platform Tokopedia dan sebanyak **89** UMKM yang menggunakan promo, detil seperti pada **Gambar 6.4**.



Gambar 6. 4 Grafik perbandingan jumlah UMKM pengguna promo

Keseluruhan data bisa dilihat pada **LAMPIRAN B**. Hasil ringkasan data dapat dilihat pada **Gambar 6.5**.

	link
0	https://www.tokopedia.com/kenriznie
1	https://www.tokopedia.com/amaliaparfume
2	https://www.tokopedia.com/infinimart
3	https://www.tokopedia.com/rizkymax
4	https://www.tokopedia.com/nona-muslimah
5	https://www.tokopedia.com/inorimala
...	...
11170	https://www.tokopedia.com/dapurrr153
11171	https://www.tokopedia.com/dewisetiyaningsi
11172	https://www.tokopedia.com/tiahome
11173	https://www.tokopedia.com/dlinshop
11174	https://www.tokopedia.com/rojabshop
11175	https://www.tokopedia.com/fruityduty
11176 rows x 1 columns	

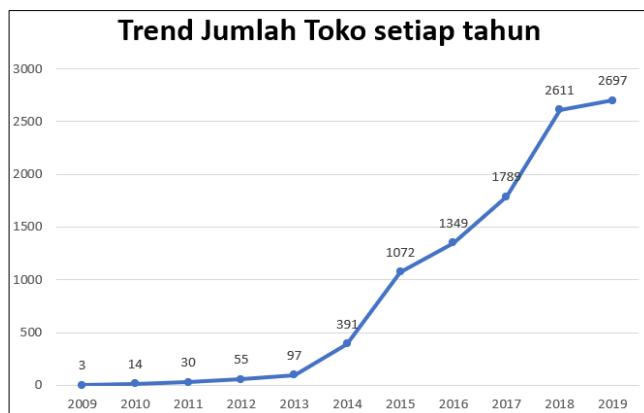
Gambar 6. 5 Data URL UMKM pada Tokopedia

d. Data UMKM pada Tokopedia

Dari hasil data dilakukan *preprocessing* sehingga siap untuk dianalisis. Terdapat **10723** informasi UMKM

Surabaya kategori *fashion muslim* yang didapatkan. Secara umum fakta pada data tokopedia kategori fashion muslim adalah sebagai berikut :

1. UMKM yang berjenis *power merchant* memiliki jumlah paling banyak yaitu **10059**, diikuti regular sebanyak **645** dan *official store* sebanyak **11** UMKM
2. UMKM memiliki rata-rata **130** *followers*, *rating* **4.5**, dan *review* sebanyak **258**
3. Pada *trend* jumlah UMKM terdapat terjadi secara kumulatif meningkat setiap tahun nya, detil dapat dilihat pada **Gambar 6.6**



Gambar 6.6 Grafik tren UMKM setiap tahun mulai dari 2009-2019

4. Pada kecepatan membala sebanyak 9563 UMKM, tidak terdeteksi dan diikuti kecepatan membala secara cepat sebanyak **570**, sangat cepat **408**, **160** sedang dan **5** lambat
5. Rata-rata produk terjual dan reputation_point masing-masing sebanyak **3818** dan **536**

Ringkasan data dan keseluruhan kolom masing-masing dapat dilihat pada **Gambar 6.7** dan **LAMPIRAN B**

	link	seller_type	...	active_product	success_transaction
0	https://www.tokopedia.com/kenriznie	Power Merchant	...	71.0	2.0
1	https://www.tokopedia.com/amaliaparfume	Power Merchant	...	14.0	3.0
2	https://www.tokopedia.com/infinimart	Power Merchant	...	36.0	2708.0
3	https://www.tokopedia.com/rizkymax	Power Merchant	...	51.0	361.0
4	https://www.tokopedia.com/nona-muslimah	Power Merchant	...	24.0	0.0
5	https://www.tokopedia.com/normalala	Power Merchant	...	137.0	331.0
...
11170	https://www.tokopedia.com/dapurrr153	Power Merchant	...	155.0	3.0
11171	https://www.tokopedia.com/dewiseityaningsi	Power Merchant	...	66.0	0.0
11172	https://www.tokopedia.com/tlahome	Power Merchant	...	114.0	0.0
11173	https://www.tokopedia.com/dlinshop	Power Merchant	...	1051.0	5323.0
11174	https://www.tokopedia.com/rojabshop	Power Merchant	...	468.0	64.0
11175	https://www.tokopedia.com/fruityduty	Power Merchant	...	353.0	1.0

11176 rows × 22 columns

Gambar 6. 7 Data UMKM pada Tokopedia

2. Hasil Data Shopee

Hasil data seperti data kategori, informasi produk, URL UMKM dan data UMKM shopee akan dibahas pada bagian ini.

a. Data Kategori Shopee

Terdapat 11 kategori fesyen muslim di shopee. Data terdiri dari nama kategori dan URL kategori. Ringkasan data dapat dilihat pada **Gambar 6.8**

	kategori	link
0	fashion muslim	https://shopee.co.id/Fashion-Muslim-cat.43
1	hijab	https://shopee.co.id/Hijab-cat.43.1085
2	atasan muslim wanita	https://shopee.co.id/Atasan-Muslim-Wanita-cat....
3	bawahan muslim wanita	https://shopee.co.id/Bawahan-Muslim-Wanita-cat...
4	dress muslim	https://shopee.co.id/Dress-Muslim-cat.43.1071
5	atasan muslim pria	https://shopee.co.id/Atasan-Muslim-Pria-cat.43...
6	bawahan muslim pria	https://shopee.co.id/Bawahan-Muslim-Pria-cat.4...
7	outerwear	https://shopee.co.id/Outerwear-cat.43.12428
8	mukena	https://shopee.co.id/Mukena-cat.43.1107
9	pakaian muslim anak	https://shopee.co.id/Pakaian-Muslim-Anak-cat.4...
10	aksesoris muslim	https://shopee.co.id/Aksesoris-Muslim-cat.43.1...
11	peralatan sholat	https://shopee.co.id/Peralatan-Sholat-cat.43.1127

Gambar 6.8 Data Kategori Shopee

b. Data Produk Shopee

Untuk mendapatkan data produk diperlukan URL kategori untuk dapat melihat data-data produk setiap kategori fesyen muslim. Didapatkan sebanyak **7619** produk pada kategori fesyen muslim.

Kemudian dilakukan pembuatan *wordcloud* untuk mengetahui produk yang paling banyak banyak dijual pada di shopee. Didapatkan kata yang paling besar muncul adalah **kaos kaki, gamis anak, gamis dan grosir**. Detil dapat dilihat pada **Gambar 6.9**



Gambar 6.9 wordcloud Shopee

Ringkasan data dapat dilihat pada **Gambar 6.10**, data secara keseluruhan dapat dilihat pada **LAMPIRAN B**

kategori	title	link
0 fashion muslim	Bella Square	https://shopee.co.id/Bella-Square-i.504653292...
1 fashion muslim	Pashmina Ceruty Babydoll Export	https://shopee.co.id/Pashmina-Ceruty-Babyd...
2 fashion muslim	41 Warna Pashmina Ceruty Import Baby Doll Pre... ...	https://shopee.co.id/41-Warna-Pashmina-Ceruty-...
3 fashion muslim	Pasma Diamond Tali BEST SELLER nissa sabyan ...	https://shopee.co.id/Pasma-Diamond-Tali-BEST-...
4 fashion muslim	VOAL MIRACEL PLAIN LASER CUT UHAMAMA 45 warna	https://shopee.co.id/VOAL-MIRACEL-PLAIN-LASER-...
5 fashion muslim	43 WARNA 17x75CM Pashmina Diamond Italiano Ge... ...	https://shopee.co.id/43-WARNA-17x75CM-Pashmin...
6 fashion muslim	PASMINA CERUTY BABYDOLL PART 1 37 WARNA PASMIN... ...	https://shopee.co.id/PASMINA-CERUTY-BABYDOLL-P...
7613 peralatan sholat	Peci Banglaes Original 100%... 1 kodi(20 pcs p... ...	https://shopee.co.id/Peci-Banglaes-Original-1...
7614 peralatan sholat	Tasbih galih kamboja 8 mm ...	https://shopee.co.id/Tasbih-galih-kamboja-8-mm-...
7615 peralatan sholat	Tasbih cendana kupang 9 mm ...	https://shopee.co.id/Tasbih-cendana-kupang-9-mm-...
7616 peralatan sholat	Tasbih Kaukah Lampu Disco ...	https://shopee.co.id/Tasbih-Kaukah-Lampu-Disco-...
7617 peralatan sholat	Tasbih 99 Kayu Galih Nagasari Alami - 10 mm ...	https://shopee.co.id/Tasbih-99-Kayu-Galih-Naga...
7618 peralatan sholat	Tasbih 33 Gaharu Wangi Asli Alami - #06 ...	https://shopee.co.id/Tasbih-33-Kayu-Gaharu-Wan...
7619 peralatan sholat	Tasbih 33 Gaharu Wangi Natural - #04 ...	https://shopee.co.id/Tasbih-33-Kayu-Gaharu-Wan...

Gambar 6. 10 Data Produk Shopee

c. Data URL UMKM Shopee

Untuk mendapatkan URL UMKM diperlukan URL produk untuk masuk ke setiap halaman produk. Terdapat **7569** data URL UMKM sesuai pada **Gambar 6.11**.

	kategori	link
0	fashion muslim	https://shopee.co.id/tazkiyah_hijab
1	fashion muslim	https://shopee.co.id/tazkiyah_hijab
2	fashion muslim	https://shopee.co.id/jasminejilbab
3	fashion muslim	https://shopee.co.id/frandashop
4	fashion muslim	https://shopee.co.id/shifasurabaya1808
5	fashion muslim	https://shopee.co.id/jasminejilbab
6	fashion muslim	https://shopee.co.id/frandashop
...
7563	peralatan sholat	https://shopee.co.id/fathimahku
7564	peralatan sholat	https://shopee.co.id/anekakayubertuah
7565	peralatan sholat	https://shopee.co.id/arunagarwoodhandicraft
7566	peralatan sholat	https://shopee.co.id/uilumilia
7567	peralatan sholat	https://shopee.co.id/nazbee
7568	peralatan sholat	https://shopee.co.id/nazbee
7569	peralatan sholat	https://shopee.co.id/nazbee
7570 rows × 2 columns		

Gambar 6. 11 Data URL UMKM Shopee

d. Data URL UMKM hasil *pre-processing*

Data yang didapat kemudian dilakukan *preprocessing* untuk menghilangkan duplikasi data. Didapatkan hasil bahwa jumlah UMKM yang berkategori *outerwear* memiliki jumlah paling sebanyak **1164** diikuti oleh akeseoris muslim sebanyak **581** dan pakaian muslim anak sebanyak **532**. Detil ada pada **Gambar 6.12**



Gambar 6. 12 Grafik jumlah UMKM setiap sub kategori fashion muslim

Hasil *preprocessing* mengashilkan data URL UMKM *unique* sebanyak **3469** data URL UMKM, sesuai pada **Gambar 6.13** dan data secara keseluruhan dapat dilihat pada **LAMPIRAN B**

kategori	link
0 fashion muslim	https://shopee.co.id/tazkiyah_hijab
1 fashion muslim	https://shopee.co.id/jasminejilbab
2 fashion muslim	https://shopee.co.id/frandashop
3 fashion muslim	https://shopee.co.id/shifasurabaya1808
4 fashion muslim	https://shopee.co.id/servantina18
5 fashion muslim	https://shopee.co.id/grosir.rakgantung
6 fashion muslim	https://shopee.co.id/qumijilbab
...	...
3463 peralatan sholat	https://shopee.co.id/meilinda_bahan_craft
3464 peralatan sholat	https://shopee.co.id/diatiyuniarisanti
3465 peralatan sholat	https://shopee.co.id/smartshoptv
3466 peralatan sholat	https://shopee.co.id/sairozy
3467 peralatan sholat	https://shopee.co.id/fathimahku
3468 peralatan sholat	https://shopee.co.id/anekakayubertuah
3469 peralatan sholat	https://shopee.co.id/ulumlia

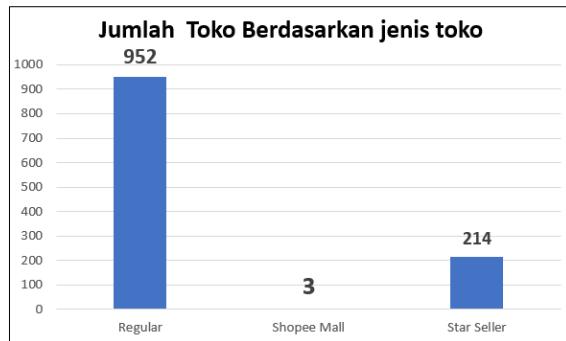
3470 rows × 2 columns

Gambar 6. 13 Data URL UMKM Hasil Pre-processing

e. Data UMKM Shopee

Data URL UMKM digunakan untuk mengakses halaman setiap UMKM, kemudian diambil seluruh informasi. Kemudian dilakukan *preprocessing* untuk agar data siap untuk dianalisis. Setelah dilakukan analisis data UMKM didapatkan sebanyak **1170** terutama UMKM di surabaya. Secara umum fakta pada data tokopedia kategori fashion muslim adalah sebagai berikut :

1. Jenis UMKM yang paling banyak adalah *Regular Seller* dengan total UMKM **952** UMKM, diikuti *Star Seller* sebanyak **214**, dan yang paling kecil adalah shopee mall hanya **3 UMKM**. Detil seperti pada **Gambar 6.14**



Gambar 6.14 Grafik jumlah UMKM berdasarkan jenis UMKM

2. Jumlah UMKM yang menyediakan *voucher* sebanyak **1152** UMKM dan UMKM yang tidak menyediakan *voucher* sebanyak **17** UMKM. Detil dapat dilihat pada **Gambar 6.15**



Gambar 6.15 Grafik perbandingan jumlah UMKM

3. Sebanyak **1152** UMKM yang tidak menyediakan sistem COD dan **17** UMKM menyediakan sistem COD
4. Rata-rata *followers* sebanyak **3829**
5. *Rating* memiliki rata-rata sebesar **4,67**.

Ringkasan data dapat dilihat pada **Gambar 6.16**, data secara keseluruhan dapat dilihat pada **LAMPIRAN B**

	name	link	...	numOf_media	description
0	ProdusenJilbab	https://shopee.co.id/tazkiyah_jilbab ...		2.0	ProdusenJilbab\nProduct-product kami Original ...
1	JasmineJilbab	https://shopee.co.id/jasminejilbab ...		3.0	JasmineJilbab\nIG : @jasminejilbab_store\nTOKO...
2	khay scarf	https://shopee.co.id/frandishop ...		4.0	khay scarf\n SELAMAT DATANG DI PRODUSEN HIJAB...
3	shifasurabaya1808	https://shopee.co.id/shifasurabaya1808 ...		3.0	shifasurabaya1808\nSELAMAT DATANG DI TOKO SHIF...
4	Shahihjab	https://shopee.co.id/servantina18 ...		3.0	Shahihjab\nBarang selalu kami kirim dalam ke...
5	Grosir.rakgantung	https://shopee.co.id/grosir.rakgantung ...		1.0	Grosir.rakgantung\nWOW PROMO BANJIR DISKON DAR...
6	QUMIJILBAB	https://shopee.co.id/qumijilbab ...		0.0	NaN
...
3463	meilinda_bahan_craft	https://shopee.co.id/meilinda_bahan_craft ...		5.0	meilinda_bahan_craft\nharap cantumkan keterangan...
3464	Dy_Boutique	https://shopee.co.id/dialityuniarisanli ...		1.0	Dy_Boutique\nSemua barang yang di toko READY...
3465	Smartsoprtv	https://shopee.co.id/smartsoprtv ...		0.0	NaN
3466	sairozy	https://shopee.co.id/sairozy ...		4.0	sairozy\nkami menjual berbagai barang drgn ku...
3467	fathimahku	https://shopee.co.id/fathimahku ...		5.0	fathimahku\nAssalamualaikum semuanya..\\sebe...
3468	anekakayubertuah	https://shopee.co.id/anekakayubertuah ...		1.0	anekakayubertuah\nKami adalah TMH yang sudah b...
3469	el_wass shop	https://shopee.co.id/elumilia ...		0.0	el_wass shop\nAulia

3470 rows × 13 columns

Gambar 6. 16 Data UMKM Shopee

6.1.2 Hasil Data Google Business

Hasil data pada proses *scrapping* Google Business dilakukan untuk mengumpulkan data kategori dan data kelurahan/kecamatan, memilih list kategori yang relevan dengan kategori *fashion muslim*, serta *scrapping* mengenai informasi UMKM pada Google Business.

a. Data kecamatan dan kelurahan

Data kecamatan didapatkan dari internet sebanyak **153** kelurahan dan kecamatan. Ringkasan data dapat dilihat pada **Gambar 17**, data secara keseluruhan dapat dilihat pada **LAMPIRAN C**

	kecamatan	kelurahan
0	Kecamatan Bubutan	Kelurahan Alun-Alun Contong
1	Kecamatan Bubutan	Kelurahan Bubutan
2	Kecamatan Bubutan	Kelurahan Gundih
3	Kecamatan Bubutan	Kelurahan Jepara
4	Kecamatan Bubutan	Kelurahan Tembok Dukuh
5	Kecamatan Genteng	Kelurahan Embong Kalasin
6	Kecamatan Genteng	Kelurahan Genteng
...
147	Kecamatan Sukomanunggal	Kelurahan Tanjungsari
148	Kecamatan Tandes	Kelurahan Balongsari
149	Kecamatan Tandes	Kelurahan Banjar Sugihan
150	Kecamatan Tandes	Kelurahan Karangpoh
151	Kecamatan Tandes	Kelurahan Manukan Kulon
152	Kecamatan Tandes	Kelurahan Manukan Wetan
153	Kecamatan Tandes	Kelurahan Tandes

154 rows × 2 columns

Gambar 6. 17 Data Kecamatan dan Kelurahan

b. Data kategori

Hasil pencarian data kategori google business, didapatkan sebanyak **3941** kategori. Ringkasan data dapat dilihat pada **Gambar 6.18**, data secara keseluruhan dapat dilihat pada **LAMPIRAN C**

	Category Name	Category ID
0	Abbey	abbey
1	Aboriginal and Torres Strait Islander organisation	aboriginal_and_torres_strait_islander_organisation
2	Aboriginal art gallery	aboriginal_art_gallery
3	Abortion clinic	abortion_clinic
4	Abrasives supplier	abrasives_supplier
5	Abundant Life church	abundant_life_church
6	Accountant	accountant
...
3935	Youth hostel	youth_hostel
3936	Youth organization	youth_organization
3937	Youth social services organization	youth_social_services_organization
3938	Yucatan restaurant	yucatan_restaurant
3939	Zac	zac
3940	Zhejiang restaurant	zhe_jiang_restaurant
3941	Zoo	zoo

3942 rows × 2 columns

Gambar 6. 18 Data Kategori

c. Data kategori hasil *pre-processing*

Hasil pemilihan kategori sesuai kategori fashion muslim, menyisakan data kategori yang relevan sebanyak **28 kategori**. Ringkasan data dapat dilihat pada **Gambar 6.19**, data secara keseluruhan dapat dilihat pada **LAMPIRAN C**

	kategori
0	baby clothing store
1	beach clothing store
2	children's clothing store
3	clothes and fabric manufacturer
4	clothes and fabric wholesaler
5	clothes market
6	clothing alteration service
...	...
22	protective clothing supplier
23	used clothing store
24	vintage clothing store
25	wedding dress rental service
26	women's clothing store
27	work clothes store
28	youth clothing store

29 rows × 1 columns

Gambar 6. 19 Data Kategori Hasil *pre-processing*

d. Data UMKM Google Business

Penggabungan data kecamatan dan data kategori yang relevan menghasilkan *keyword* untuk melakukan pencarian pada UMKM. Format *keyword* adalah ‘*data kecamatan/kelurahan + kategori*’. Informasi seperti nama, jam buka, kategori, alamat, *rating* dan *review* akan di *scrapping*. Terdapat **7920**. Secara umum berikut fakta-fakta data dari data UMKM google business :

- a. Rata-rata jam buka UMKM di seluruh Surabaya adalah **49 jam** selama seminggu. Dengan spesifikasi jika dilihat dari kecamatan, rata-rata jam buka UMKM terlama berapa pada kecamatan **genteng** yaitu **63 jam** dan yang paling kecil ada di kecamatan **asemrowo** yaitu selama **37 jam**
- b. Kategori **clothing store** memiliki jumlah UMKM paling banyak yaitu **3619** dan yang paling rendah ada pada kategori **dress and tuxedo rental service, full dress rental services, plus size clothing store** dan **protective clothing supplier** sebanyak **1**

Secara umum ringkasan data dan data keseluruhan masing -masing dapat dilihat pada **Gambar 6.20** dan **LAMPIRAN B**

		name	website	...	phone	instagram
0		Target	https://www.target.com/sl/culpeper/2250 ...	(540) 829-8613	NaN	
1		Green Bean Baby Boutique	https://www.greenbeanbabyboutique.com/ ...	(434) 234-4504	NaN	
2		Target	https://www.target.com/sl/front-royal/2297 ...	(540) 631-3290	NaN	
3		Walgreens	https://www.walgreens.com/locator/walgreens-59... ...	(540) 727-0483	NaN	
4	Kohl's Culpeper- Curbside Pickup only		https://www.kohls.com/stores/va/culpeper-1062... ...	(540) 829-0531	NaN	
5		Walmart Supercenter	https://www.walmart.com/store/2564/luray-va/de... ...	(540) 743-4111	NaN	
6		Walgreens	https://www.walgreens.com/locator/walgreens-12... ...	(540) 459-2183	NaN	
...	
84705	DICK'S Sporting Goods		https://stores.dicks sporting goods.com/va/culpe... ...	(540) 317-4839	NaN	
84706	Kohl's Culpeper- Curbside Pickup only		https://www.kohls.com/stores/va/culpeper-1062... ...	(540) 829-0531	NaN	
84707		Target	https://www.target.com/sl/front-royal/2297 ...	(540) 631-3290	NaN	
84708		Walmart Supercenter	https://www.walmart.com/store/2564/luray-va/de... ...	(540) 743-4111	NaN	
84709		Walmart Supercenter	https://www.walmart.com/store/2136/culpeper-va... ...	(540) 826-2723	NaN	
84710		Walmart Supercenter	https://www.walmart.com/store/2437/warrenton-v... ...	(540) 341-3568	NaN	
84711		Walmart Supercenter	https://www.walmart.com/store/5105/front-royal... ...	(540) 635-4140	NaN	

84712 rows × 10 columns

Gambar 6. 20 Data UMKM Google Business

6.1.3 Hasil Data Instagram

Hasil *data scrapping* di Instagram dilakukan pada 2 platform yaitu tokopedia dan shopee.

a. Tokopedia

Pada hasil *data scrapping* menggunakan nama UMKM di Tokopedia didapatkan **214.344** baris. Hasil masih terdiri dari pencarian ‘#’, *location* dan *explore*. Ringkasan hasil data dapat dilihat pada **Gambar 6.21** dan kode lengkap bisa dilihat di **LAMPIRAN B**.

	store	city	account	description	link
0	Nirwani hijab sby	Kota Surabaya	nirwani_hijab_sby	Nirwani hijab sby	https://www.instagram.com/nirwani_hijab_sby/
1	Nirwani hijab sby	Kota Surabaya	Nirwani hijab sby	Jl kalianget , Kota Surabaya	https://www.instagram.com/explore/locations/37...
2	amalia perfume	Kota Surabaya	amalialperfumes	AmaliaPerfumes	https://www.instagram.com/amalialperfumes/
3	amalia perfume	Kota Surabaya	amalialperfume	Amalia Perfumes	https://www.instagram.com/amalialperfume/
4	amalia perfume	Kota Surabaya	amalia_perfume	Amalia_perfume	https://www.instagram.com/amalia_perfume/
...
214339	Fruity duty	Tenggiling Mejoyo	_fart_poop_duty_	_secretly_fruity_	https://www.instagram.com/_fart_poop_duty_/
214340	Fruity duty	Tenggiling Mejoyo	fruitydutyrouly	Karina johns	https://www.instagram.com/fruitydutyrouly/
214341	Fruity duty	Tenggiling Mejoyo	dutifruity	NaN	https://www.instagram.com/dutifruity/
214342	Fruity duty	Tenggiling Mejoyo	fruitymcdufty	Chris Dial Ass	https://www.instagram.com/fruitymcdufty/
214343	Fruity duty	Tenggiling Mejoyo	hdf.ina	heavy duty fruity	https://www.instagram.com/hdf.ina/

214344 rows x 5 columns

Gambar 6. 21 Hasil data scrapping instagram

b. Shopee

Hasil *data scrapping* menggunakan nama UMKM pada shopee didapatkan **41.514** data. Data masih belum di *preprocess* sehingga masih terdapat hasil pencarian seperti ‘#’, *location* dan *explore*. Ringkasan hasil data dapat dilihat pada **Gambar 6.22**. Keseluruhan data terdapat pada **LAMPIRAN B**

	store	city	account	description	link
0	ProdusenJilbab	KAB. GRESIK	#produsenjilbab	160,492 posts	https://www.instagram.com/explore/tags/produse...
1	ProdusenJilbab	KAB. GRESIK	produsenjilbab_solo	produsen & grosir jilbab	https://www.instagram.com/produsenjilbab_solo/
2	ProdusenJilbab	KAB. GRESIK	produsenjilbabupdate	Hijab Grosir Surabaya	https://www.instagram.com/produsenjilbabupd...
3	ProdusenJilbab	KAB. GRESIK	rahilihjab_on	PRODUSEN JILBAB KAOS PREMIUM	https://www.instagram.com/corrahilihjab_on/
4	ProdusenJilbab	KAB. GRESIK	_purplerainbow	purple rainbow	https://www.instagram.com/_purplerainbow/
...
41509	fathimahku	KAB. MAGETAN	fathimahkuliner	fathimah	https://www.instagram.com/fathimahkuliner/
41510	fathimahku	KAB. MAGETAN	fathimah_kukur	newbie ☺	https://www.instagram.com/fathimah_kukur/
41511	fathimahku	KAB. MAGETAN	fathimahkusuma	Fathimah Kusuma	https://www.instagram.com/fathimahkusuma/
41512	anekakayuberliah	KAB. JEMBER	aneka_kayu_berliah	aksesoris gelang antik	https://www.instagram.com/aneka_kayu_berliah/
41513	el_wass shop	KAB. BANYUWANGI	el_wass_shop	Abu Muhammad ☺	https://www.instagram.com/el_wass_shop/

Gambar 6. 22 Hasil data scrapping shopee

6.2 Hasil Data Pre-processing

Tahap *pre-processing* dilakukan untuk menjamin data bersih serta siap digunakan. Terdapat beberapa pengujian pada tahap *pre-processing*, yaitu antara lain: *data cleaning*, *data transformation* dan *data integration..* Berikut hasil *pre-processing* yang dilakukan pada Tokopedia, Shopee, Google Business dan Instagram.

6.2.1 Hasil Data Pre-Processing Tokopedia

Beberapa kegiatan data *pre-processing* pada Tokopedia seperti *data cleaning* dan *data transformation*. Menyisakan **10722** baris UMKM Tokopedia sesuai dengan **Gambar 6.23**. Hasil data *preprocessing* bisa dilihat pada **LAMPIRAN B**

	name	link	...	website	instagram
0	Nirwani hijab sby	https://www.tokopedia.com/kenriznie	...	no website	no instagram
1	amalia perfume	https://www.tokopedia.com/amalliaperfume	...	no website	no instagram
2	Infinimart	https://www.tokopedia.com/infinimart	...	no website	no instagram
3	RizkyMax	https://www.tokopedia.com/rizkymax	...	no website	no instagram
4	Nona_Muslimah	https://www.tokopedia.com/nona-muslimah	...	no website	no instagram
5	inorimala	https://www.tokopedia.com/inorimala	...	no website	no instagram
6	Aksesoris Pengantin	https://www.tokopedia.com/aksesorispesta	...	no website	no instagram
...
10716	Fendy Olshop86	https://www.tokopedia.com/fendyolshop86	...	no website	no instagram
10717	Dapurrr 153	https://www.tokopedia.com/dapurrr153	...	no website	no instagram
10718	DEWI SETIYANINGSIH	https://www.tokopedia.com/dewisetiyaningsi	...	no website	no instagram
10719	Tiahome	https://www.tokopedia.com/tiahome	...	no website	no instagram
10720	Diin Shop	https://www.tokopedia.com/diinshop	...	no website	no instagram
10721	rojab shop	https://www.tokopedia.com/rojabshop	...	no website	no instagram
10722	Fruity duty	https://www.tokopedia.com/fruityduty	...	no website	no instagram

10723 rows × 30 columns

Gambar 6. 23 Hasil Data Pre-Processing Tokopedia

6.2.2 Hasil Data Pre-Processing Shopee

Seperti halnya Tokopedia, beberapa kegiatan data *pre-processing* pada shopee seperti *data cleaning* dan *data transformation*. Menyisakan **1170** baris UMKM shopee sesuai dengan **Gambar 6.24**. Keseluruhan data bisa dilihat pada **LAMPIRAN B**

	name	link	...	website	instagram
0	shifasurabaya1808	https://shopee.co.id/shifasurabaya1808	...	no website	no instagram
1	greenday.veil	https://shopee.co.id/greenday.veil	...	no website	greey.hijab
2	quindiragrosir	https://shopee.co.id/quindiragrosir	...	no website	no instagram
3	Dianhijabsurabaya	https://shopee.co.id/dianhijabsurabaya	...	no website	no instagram
4	kus_14	https://shopee.co.id/kus_14	...	no website	no instagram
5	hijabmuslimahcantix	https://shopee.co.id/hijabmuslimahcantix	...	no website	no instagram
6	TrFashionStory	https://shopee.co.id/tichazh	...	no website	no instagram
...
1163	Antonio_Shop_Baju	https://shopee.co.id/antonio_shop_baju	...	no website	no instagram
1164	anitaptw47	https://shopee.co.id/anitaptw47	...	no website	no instagram
1165	marianaangel	https://shopee.co.id/marianaangel	...	no website	no instagram
1166	berlianpacific	https://shopee.co.id/berlianpacific	...	no website	no instagram
1167	meilinda_bahan_craft	https://shopee.co.id/meilinda_bahan_craft	...	no website	no instagram
1168	Smartshoptv	https://shopee.co.id/smartshtoptv	...	no website	no instagram
1169	sairozy	https://shopee.co.id/sairozy	...	no website	no instagram
1170 rows × 19 columns					

Gambar 6. 24 Ringkasan data UMKM pada data Pre-Processing Shopee

6.2.3 Hasil Data Pre-Processing Google Business

Kegiatan *preprocessing* pada google business juga dilakukan seperti mengkonversi atribut *operational_hours* menjadi *operational_hours_duration* agar menjadi durasi yang dapat dihitung, dengan implementasi lengkap seperti pada bagian. Terdapat **7919** hasil UMKM yang telah dilakukan

preprocessing, detil pada **Gambar 6.25**. Data lengkap bisa dilihat pada **LAMPIRAN B**

	name	website	...	tokopedia	shopee
0	kimmy baby dan kids shop	no website	...	tokopedia not found	shopee not found
1	kimmie baby shop	no website	...	tokopedia not found	shopee not found
2	ud adi mitra	no website	...	tokopedia not found	shopee not found
3	toko endang	no website	...	tokopedia not found	shopee not found
4	toko sundari	no website	...	tokopedia not found	shopee not found
5	toko sinar sentosa	no website	...	tokopedia not found	shopee not found
6	toko pak rahmat	no website	...	tokopedia not found	shopee not found
...
7913	toko beautifull kids	https://toko-beautifull-kids.business.site/	...	tokopedia not found	shopee not found
7914	toko ghalang kids	https://toko-ghalang-kids.business.site/	...	tokopedia not found	shopee not found
7915	pom pom kids	no website	...	tokopedia not found	shopee not found
7916	mahkota kids - pusat grosir surabaya	no website	...	tokopedia not found	shopee not found
7917	centro kids	no website	...	tokopedia not found	shopee not found
7918	shopp	no website	...	tokopedia not found	shopee not found
7919	grosir kaos distro surabaya	no website	...	tokopedia not found	shopee not found
7920 rows × 15 columns					

Gambar 6. 25 Hasil Data Pre-Processing Google Business

6.2.4 Hasil Data Pre-Processing Instagram

a. Tokopedia – Instagram

Hasil *pre-processing* data Instagram yaitu dengan menghilangkan pencarian yang mengandung *#*, *location*, dan *explore* untuk mendapatkan akun instagram. Didapatkan total **110.290 akun**. Ringkasan data bisa dilihat pada **Gambar 6.26**. Keseluruhan data bisa dilihat pada **LAMPIRAN B**

	store	city	account	description	link
0	nirwani hijab sby	kota surabaya	nirwani_hijab_sby	Nirwani hijab sby	https://www.instagram.com/nirwani_hijab_sby/
1	nirwani hijab sby	kota surabaya	Nirwani Hijab sby	Jl kalianget, Kota Surabaya	https://www.instagram.com/explore/locations/37...
2	amalia perfume	kota surabaya	amatlaperfumes	AmaliaPerfumes	https://www.instagram.com/amatlaperfumes/
3	amalia perfume	kota surabaya	amatlaperfume	Amalia Perfumes	https://www.instagram.com/amatlaperfume/
4	amalia perfume	kota surabaya	amalia_perfume	Amalia_Perfume	https://www.instagram.com/amalia_perfume/
...
110286	din shop	kota surabaya	dingyshop	dingy shop	https://www.instagram.com/dingyshop/
110287	din shop	kota surabaya	dinshop90	Sari P Dinni	https://www.instagram.com/dinshop90/
110288	din shop	kota surabaya	dinanshop_	ONLINE SHOP KENDARI	https://www.instagram.com/dinanshop_/
110289	din shop	kota surabaya	Dinshop16	Dinshop & Jastip	https://www.instagram.com/dinshop16/
110290	din shop	kota surabaya	dinar_shop	dinar shop	https://www.instagram.com/dimar_shop/
110291 rows × 5 columns					

Gambar 6. 26 Data Pre-Processing Instagram

b. Shopee – Instagram

Hasil *pre-processing* data shopee dan instagram menghasilkan total 9404 data akun instagram. Detil ringkasan data bisa dilihat pada **Gambar 6.27**. Data keseluruhan bisa dilihat pada **LAMPIRAN B**

	store	city	account	description	link
0	shifasurabaya1908	kota surabaya	NaN	NaN	NaN
1	greenday veil	kota surabaya	chemical_black_veil_greenday	my_black_chemical_greenbrides	https://www.instagram.com/chemical_black_veil_...
2	greenday veil	kota surabaya	blackveilbrides_greenday	blackveilbrides_greenday	https://www.instagram.com/blackveilbrides_gree...
3	greenday veil	kota surabaya	greenday_bvb	Green Day & Black Veil Brides	https://www.instagram.com/greenday_bvb/
4	greenday veil	kota surabaya	live_properly_eat_chocolate	PIERCE THE VEIL ~ GREENDAY	https://www.instagram.com/live_properly_eat_ch...
...
5399	sairozy	kota surabaya	sairozsairoz7085	SAIROZI SAIROZI	https://www.instagram.com/sairozsairoz7085/
9400	sairozy	kota surabaya	sairozi sairozi 3	Sairozi	https://www.instagram.com/sairozi_sairozi_3/
9401	sairozy	kota surabaya	sairoz1980	Mohamad Sairozi Rashedi	https://www.instagram.com/sairoz1980/
9402	sairozy	kota surabaya	sairozidasht	NaN	https://www.instagram.com/sairozidasht/
9403	sairozy	kota surabaya	m.sairozi	sairozi	https://www.instagram.com/m.sairozi/

Gambar 6. 27 Data Pre-Processing Shopee Instagram

6.2.5 Hasil Data Integration

Kegiatan *data integration*, menggunakan metode *entity resolution* yaitu dengan nilai similaritas atau kesamaan dari suatu atribut data menggunakan library *sequencematcher*. Dipilih atribut data yang sama yaitu atribut *phone*, *website*, *instagram* dan *name*. Terdapat 2 hasil data integrasi dengan tujuan untuk melihat UMKM secara *online*, *offline*, dan *offline/online*. Berikut penjelasan kedua hasil data integrasi :

a. Data Tokopedia, Shopee & Instagram

Data pertama didapat dari hasil integrasi data tokopedia, shopee dan similaritas instagram, dengan secara umum integrasi digunakan untuk melihat kondisi UMKM secara

online. Hasil lengkap data bisa dilihat pada **LAMPIRAN**
B. Ringkasan data bisa dilihat pada **Gambar 6.28**

name_tokopedia	link_tokopedia	seller_type_tokopedia	...	instagram_shopee	shopee_lg_merged
Nirwani hijab sby	https://www.tokopedia.com/kenriznie	Power Merchant	...	NaN	NaN
amalia perfume	https://www.tokopedia.com/amaliaparfume	Power Merchant	...	NaN	NaN
Infinimart	https://www.tokopedia.com/infinimart	Power Merchant	...	NaN	NaN
RizkyMax	https://www.tokopedia.com/rizkymax	Power Merchant	...	NaN	NaN
Nona_Muslimah	https://www.tokopedia.com/nona-muslimah	Power Merchant	...	NaN	NaN
...
NaN	NaN	NaN	...	no phone	no instagram
NaN	NaN	NaN	...	no phone	no instagram
NaN	NaN	NaN	...	no phone	no instagram
NaN	NaN	NaN	...	no phone	no instagram
NaN	NaN	NaN	...	no phone	no instagram

10731 rows × 53 columns

Gambar 6. 28 Hasil Data Integration Tokopedia, shopee dan instagram

b. Data Google My Business,Tokopedia, instagram dan Shopee

Data kedua, menggunakan data google business, tokopedia dan shopee untuk melihat kondisi secara *offline* dan *offline/online*, Hasil lengkap data bisa dilihat pada

LAMPIRAN B. Ringkasan data bisa dilihat pada Gambar 6.29

index		name	website	rating	...	tokopedia_similarity_by	shopee_similarity	shopee_similarity_by
0		kimmy baby dan kids shop	website not found	NaN	...	match not found	0.0	match not found
1		kimmie baby shop	website not found	5.0	...	match not found	0.0	match not found
2		ud adi mitra	website not found	4.0	...	match not found	0.0	match not found
3		toko endang	website not found	NaN	...	match not found	0.0	match not found
4		toko sundari	website not found	NaN	...	match not found	0.0	match not found
...	
7929		pom pom kids	website not found	NaN	...	match not found	0.0	match not found
7930		mahkota kids - pusat grosir surabaya	website not found	NaN	...	match not found	0.0	match not found
7931		centro kids	website not found	NaN	...	match not found	0.0	match not found
7932		shopp	website not found	NaN	...	match not found	0.0	match not found
7933		grosir kaos distro surabaya	website not found	5.0	...	match not found	0.0	match not found

7934 rows × 19 columns

Gambar 6. 29 Hasil Data Integration Google My Business,Tokopedia dan Shopee

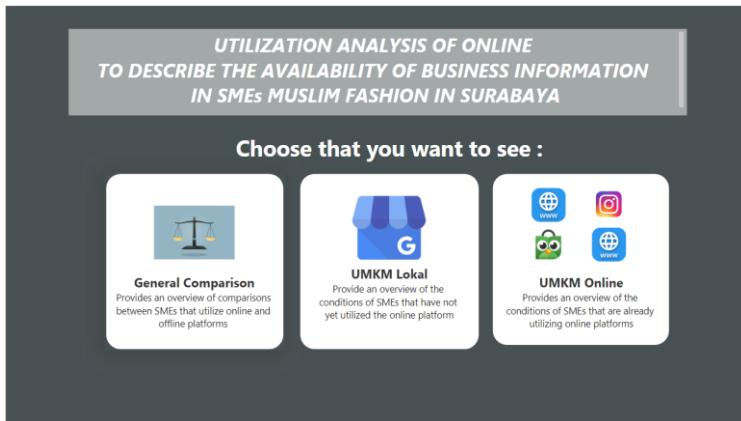
6.3 Data Validation

Dari hasil hasil implementasi bagian 5.3 Didapatkan hasil bahwa nilai rata-rata *precision* hasil integrasi pada penggunaan dua metode yang berbeda yaitu sebesar **61%** untuk metode *sequencematcher* dan **17%** untuk metode *Intersection of Words*. Pada hasil penggunaan library *sequencematcher* proses integrasi memakan waktu lebih lama yaitu rata-rata berkisar 30 menit daripada metode *Intersection of Words* rata-rata hanya berkisar 5 menit. Namun rata-rata skor *precision* yang memiliki nilai paling besar tetap dipilih sebagai hasil yang representatif. Sehingga sebesar **61%** data hasil integrasi dikatakan *valid* sesuai data di lapangan.

6.4 Hasil Visualisasi Data dan Analisis

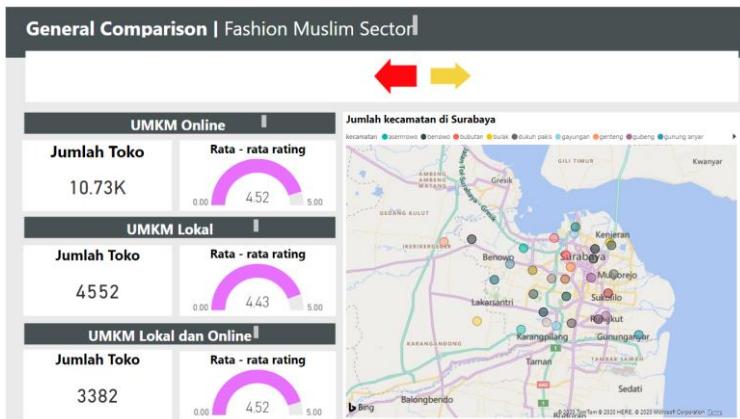
Proses visualisai data ini dilakuakn dengan cara mengelola *data loaded* yang ada pada *tools* untuk menjadi informasi visual seperti *diagram* atau *chart* sesuai hasil *metrics* visualisai yang telah dibuat. Digunakan *dashboard* sebagai *tools* untuk melakukan visualisasi data. Secara umum *dashboard* menjelaskan hasil visualisasi data berdasarkan platform *offline* dan *online*. *Dashboard* dibagi menjadi 3 menu, *General*

Comparison, offline dan online. Pada laman awal dashboard sesuai **Gambar 6.30**.



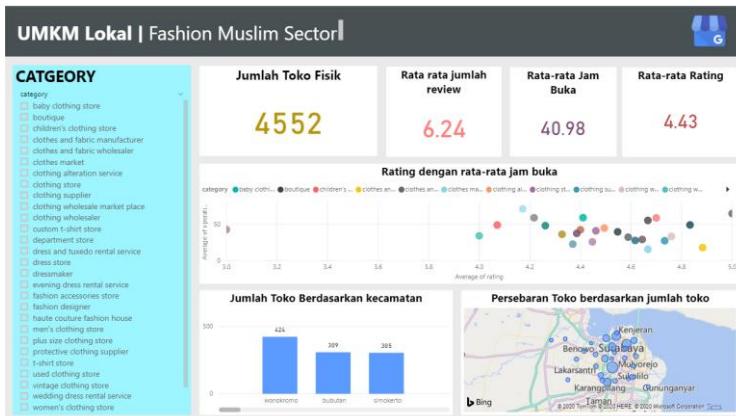
Gambar 6. 30 Laman awal Dashboard

Pada halaman *General Comparison* menjelaskan kondisi UMKM dari secara umum pada platform *online*, UMKM fisik (*offline*) dan UMKM fisik (*offline*) dan *online*. Dari visualisasi data yang ditunjukkan bahwa saat ini banyak UMKM yang memanfaatkan *platform online* tanpa penggunaan lapak fisik. Hal ini dibuktikan dengan jumlah UMKM *online* sebanyak **10 ribu UMKM**, Dibanding dengan lapak fisik hanya sebanyak **4552 UMKM**. Visualisasi *general comparison* dapat dilihat pada **Gambar 6.31**.



Gambar 6. 31 Visualisasi general comparison

Pada laman *dashboard* UMKM fisik (*offline*), ditampilkan jumlah UMKM *online* (tanpa offline) rata-rata jumlah review, rata-rata jumlah jam buka UMKM, rata-rata rating, hubungan antar jam buka, rating dan kategori serta persebaran jumlah UMKM berdasarkan kecamatan. Hasil visualisasi didapatkan informasi bahwa persebaran kategori disesuaikan dengan rating dan jam buka hasilnya merata, sehingga dapat disimpulkan jam buka yang lama tidak menjamin rating akan tinggi. Persebaran dan jumlah 3 UMKM teratas didominasi kecamatan di daerah pusat kota yaitu wonokromo, bubutan dan simokerto, sehingga selurus dengan jumlah UMKM yang juga cukup banyak. Visualisasi pada halaman UMKM fisik bisa dilihat pada **Gambar 6.32**.



Gambar 6.32 Visualisasi halaman UMKM fisik

Pada laman *dashboard* UMKM online disajikan masing-masing kondisi UMKM dari Tokopedia dan shopee pada **Gambar 6.33** dan **Gambar 6.34**. Terdapat kesamaan variabel yaitu variabel follower dan jumlah *review*. Didapatkan hasil bahwa semakin banyak jumlah *follower* maka semakin banyak jumlah *review*. Pada laman tokopedia didapatkan fakta bahwa semakin banyak produk terjual maka seiring dengan *rating* yang semakin tinggi. Tren jumlah UMKM setiap tahun juga meningkat ini menandakan bahwa masyarakat juga cenderung memilih platform Tokopedia sebagai pilihan platform jual beli secara *online*.

Pada shopee, jumlah UMKM mayoritas berjenis *regular seller* yaitu sebanyak **953 UMKM** dibanding dengan *star seller* yaitu sebanyak **214 UMKM**. Sehingga dapat disimpulkan bahwa masih banyak UMKM yang belum secara konsisten memanfaatkan fasilitas yang disediakan shopee dengan maksimal.



Gambar 6. 33 Visualisasi pada halaman UMKM Tokopedia



Gambar 6. 34 Visualisasi Pada Halaman UMKM Shopee

Halaman ini sengaja dikosongkan

BAB VI

Kesimpulan dan Saran

Bab ini membahas kesimpulan penelitian yang telah dilakukan dan saran yang diusulkan. Berikut ini adalah kesimpulan dan saran penelitian serupa di masa mendatang.

7.1 Kesimpulan

Secara umum tahap yang dilakukan dalam penelitian ini meliputi pengambilan data menggunakan metode *web scrapping*, melakukan *pre-processing* atau pengolahan terhadap data yang telah didapatkan, integrasi data, validitas kebenaran data hingga proses visualisasi informasi. Tujuan utama dari penelitian ini adalah untuk mengetahui gambaran kondisi UMKM dalam memanfaatkan berbagai macam platform *online*. Dalam melakukan penelitian, terdapat pembelajaran yang didapatkan oleh peneliti. Berikut ini pembelajaran yang didapatkan:

1. Keterbatasan akses internet, dan tidak adanya server mumpuni menjadi hambatan peneliti dalam melakukan *data scrapping* sehingga data tidak maksimal didapatkan. Tindak lanjut yang dilakukan oleh peneliti adalah menggunakan *Virtual Private Server (VPS)* agar proses *data scrapping* menjadi lebih optimal.
2. Pada proses nya seluruh data dari berbagi platform *online* (Google My Business, Tokopedia, Shopee dan Instagram) akan diintegrasikan dengan melihat kesamaan antar atribut sehingga data menjadi satu kesatuan yang dapat dianalisis. Hasil integrasi data kemudian dilakukan uji validitas untuk mengukur sejauh mana ketepatan data di lapangan dan data yang dilaporkan peneliti. Berikut ini pembelajaran yang didapatkan pada uji validitas:
 - a. Penentuan *confidence level* ditentukan sebesar 90% dengan margin error sebesar 10%. Alasan pemilihan

- adalah semakin tinggi tingkat *confidence level* maka semakin kecil nilai presisi yang akan didapatkan.
- b. Dalam melakukan uji validitas di lapangan ditemukan tantangan untuk memastikan antar platform yang diperiksa benar-benar milik UMKM sesuai hasil integrasi data. Hal-hal seperti melihat dengan teliti kesamaan nama, alamat, foto profil wajib dilakukan untuk meningkatkan keyakinan peneliti bahwa hasil benar-benar tepat (*True Positive*).

Secara umum hasil dua integrasi data menggunakan pada penitian ini memiliki tingkat validitas data sebesar **61%** dilihat dari rata-rata nilai *precision* dengan menggunakan *library sequencematcher*. Hasil validitas data dapat menjadi gambaran kondisi sejauh mana data UMKM *fashion muslim* pada platform *online* dikatakan *valid* sesuai kondisi lapangan. Pada penelitian berikutnya dapat dilakukan optimasi hasil integrasi dan validasi sehingga skor presisi dapat meningkat.

- 3. Informasi visual dilakukan dengan tujuan memberikan gambaran kondisi data UMKM kategori fesyen muslim di Surabaya dari hasil *data scraping*. Hasil analisis data dapat digunakan oleh pemerintah kota Surabaya dan peneliti sebagai referensi untuk melakukan kajian lebih lanjut. Visualisasi didapatkan dengan melakukan pengumpulan dan pengolahan data, kemudian dilakukan pemetaan metrics untuk mengetahui tujuan visualisasi berdasarkan jenis UMKM baik *online* maupun lokal. Hasil analisis data visualisasi adalah sebagai berikut:
 - a. Saat ini, sudah banyak UMKM yang memanfaatkan *platform online* dalam proses jual beli. Hal ini dibuktikan dengan besarnya jumlah UMKM *online* yaitu 10 ribu dibanding dengan UMKM fisik hanya 4552 UMKM. Dapat diartikan, secara kuantitas UMKM sudah baik dalam memanfaatkan *platform online*.

- b. Secara kualitas, pemanfaatan *platform online* oleh UMKM kategori *fashion muslim* belum cukup baik. Melihat bukti besarnya UMKM online yang berjenis *Regular Seller* dibanding *Star Seller* yang notabene label untuk UMKM yang aktif dan mampu melayani dengan baik.
- c. Pada sisi UMKM lokal secara kualitas juga masih kurang baik dalam memanfaatkan *platform online*. Terbukti dari rata-rata *rating* dan *follower* yang lebih rendah daripada UMKM yang menggunakan *platform online*.

7.2 Saran

Dalam pelaksanaan tugas akhir, terdapat beberapa saran yang diharapkan dapat bermanfaat bagi penelitian selanjutnya dan pemerintah.

- 1. Saran untuk penelitian berikutnya
 - a. Pada integrasi dan uji validitas perlu dilakukan kajian untuk mendapatkan nilai optimal dan hasil similaritas yang lebih banyak. Penggunaan banyak opsi *library* atau metode dalam melakukan integrasi dapat dilakukan untuk mendapatkan perbandingan hasil, sehingga dapat dipilih nilai yang paling optimal.
 - b. Beberapa opsi *threshold* juga perlu digunakan dalam melakukan integrasi data. Adanya opsi *threshold* dapat memudahkan peneliti berikutnya untuk dapat membandingkan hasil perolehan integrasi similaritas data sehingga dapat dipilih dan ditentukan nilai *threshold* yang paling optimal.
 - c. Perlu dilakukan analisis data secara diagnostik, preskriptif maupun prediktif pada penelitian berikutnya. Sehingga hasil analisis data dapat secara luas dan akurat didapatkan sesuai kondisi lapangan.

2. Saran untuk Pemerintah

Perlu dilakukan pengkajian lebih lanjut oleh pemerintah. Kajian data dapat dilakukan secara diagnostik, preskriptif dan prediktif dengan data yang telah didapatkan oleh peneliti untuk mendukung pengambilan keputusan ekonomi terkait dalam pengelolaan usaha, keputusan pengembangan pasar dan penetapan harga, baik secara operasional, taktis atau strategis. Pemerintah juga dapat melakukan survei dengan data yang dimiliki oleh pemerintah atau dinas kota Surabaya dengan menggabungkan data yang didapatkan oleh peneliti. Sehingga hasil analisis data dapat lebih akurat dan tepat sasaran.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- [1] K. K. d. U. K. d. M. R. Indonesia, "ARAH KEBIJAKAN BIDANG KOPERASI DAN USAHA MIKRO, KECIL DAN MENENGAH," Kementerian Koperasi dan Usaha Kecil dan Menengah Republik Indonesia, Jakarta, 2005.
- [2] B. I. d. LPPI, "PROFIL BISNIS USAHA MIKRO, KECIL DAN MENENGAH (UMKM)," Bank Indonesia dan LPPI, 2015.
- [3] D. A. Harahap, "PERILAKU BELANJA ONLINE DI INDONESIA: STUDI KASUS," *Jurnal Riset Manajemen Sains Indonesia (JRMSI)*, vol. 9, no. 2, 2018.
- [4] M. Q. Pathon, Qualitative Evaluation Methods, London: Sage Publications, 1980.
- [5] K. P. R. Indonesia, "Warta Ekspor : Fesyen Muslim Indonesia," Jakarta, 2015.
- [6] Andrio, "Analisis Inovasi Berkelanjutan pada Usaha, Mikro, Kecil dan Menengah (UMKM) Industri busana muslim di kota Bandung," Universitas Katolik Parahyangan, Bandung, 2018.
- [7] O. Castrillo-Fernández, Web Scrapping : An Application Tools, 2015.
- [8] H. S. P. H. M. Dhita Deviacita Ayani, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 4, 2019.
- [9] E. R. Hanna Kopcke, "Framework for entity matching: A comparison," University of Leipzig, Leipzig, 2009.
- [10] S. A. S. M. A. A. M. A. T. K. Jawahire Nakash*, "Real Time Product Analysis using Data Mining," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* , vol. 4, no. 3, p. 815, 2015.

- [11] A. M. S. A. D. K. S. N. B. U. S. D. Pratiba, "Web Scraping dan Data Acquisition Using Google Scholarq," in *Computational Systems and Information Technology for Sustainable Solutions*, Bangalore, 2018.
- [12] I. Primiana, "Menggerakkan Sektor Riil UKM & Industri," *Alfabeta*, Bandung, 2009.
- [13] A. L.-P. a. R. P. C. Holsapple, "A unified foundation for business analytics," *Decis. Support Syst.*, vol. 64, pp. 130-141, 2014.
- [14] M. Grigsby, *Advanced Customer Analytics: Targeting, Valuing, Segmenting and Loyalty Techniques*, Kogan Page Publishers, 2016.
- [15] S. M. S. Kabir, *Methods of data Collection*, 2016.
- [16] O. Castrillo-Fernández, *Web Scrapping: Applications an Tools*, 2015.
- [17] L. D. G. I. W. Zahraa S. Abdallah, *Encyclopedia of Machine Learning and Data Mining*, Melbourne: Springer, 2017.
- [18] H. W. J. L. H. G. Yiming Lin, "Efficient Entity Resolution on Heterogeneous Records," Harbin Institute of Technology, China, 2018.
- [19] N. F. T. G. S. G. U. G. J. P. L. Q. K. M. S. K. t. B. M. v. d. L. K. W. Marco Di Zio, *Methodology for data validation 1.0*, Essnet Validat Foundation, 2016.
- [20] Netquest, *A Comprehensive Guide to Data Visualization*, Netquest, 2018.
- [21] P. G. K. S. Jasdeep Singh Malik, "A Comprehensive Approach Towards Data Preprocessing Techniques & Association Rules," IES-IPS Academy, India.
- [22] U. Y. S. Samudera, "Analisis Pemanfaatan platform daring pada usaha mikro kecil menengah bidang makanan dan minuman di Surabaya," Surabatya, 2020.

- [23] H. C. J. L. R. W. (Eds.), Trends and Applications in Knowledge Discovery and Data Mining, Auckland: Springer, 2016.
- [24] N. Setiawan, "PENENTUAN UKURAN SAMPEL MEMAKAI RUMUS SLOVIN DAN TABEL KREJCIE-MORGAN: TELAAH KONSEP DAN APLIKASINYA," FAKULTAS PETERNAKAN UNIVERSITAS PADJADJARAN, Bandung, 2007.
- [25] M. K. Peter A. Flach, "Precision-Recall-Gain Curves: PR Analysis Done Right," University of Bristol, United Kingdom, Bristol, 2015.
- [26] A. S. D. Ramadhan, "Analisis pemanfaatan platform daring pada UMKM bidang kesehatan di Surabaya," Surabaya, 2020.

Halaman ini sengaja dikosongkan

LAMPIRAN A Kode Program

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

A. Kode *Scraping* data

Pada bagian ini dilampirkan *script* yang digunakan dalam proses *scrapping*

1. Tokopedia

Tabel A. 1 Scrapping Script Kategori Fashion Muslim Tokopedia

Kode <i>scrapping</i> kategori fashion muslim Tokopedia	
URL	https://drive.google.com/file/d/1MW9m4h4RKsoR8RbP4CMu_bCtMK9eTvX/view?usp=sharing

Tabel A. 2 Scrapping Script Produk Fashion Muslim Tokopedia

Kode <i>scrapping</i> produk fashion muslim Tokopedia	
U R L	https://drive.google.com/file/d/1xroyyweDMyUbQ4H75ji3F4xFbvB2eZS/view?usp=sharing

Tabel A. 3 Kode preprocessing URL Produk

Kode <i>preprocess URL</i> produk menjadi URL UMKM	
U R L	https://drive.google.com/file/d/159zlcMYHZjD2qKVqreEPAPzGBxLmDbzT/view?usp=sharing

Tabel A. 4 Scrapping Script Informasi Kategori Fashion Muslim Tokopedia

Kode <i>scrapping</i> informasi UMKM kategori fashion muslim Tokopedia	

UR	https://drive.google.com/file/d/1BR9WMUZXklmDb0t2YAveV-Xjl6E5ZVs /view?usp=sharing
L	

2. Shopee

Tabel A. 5 Scrapping Script Kategori Fashion Muslim Shopee

Kode scrapping kategori fashion muslim Shopee	
U R L	https://drive.google.com/file/d/12Pf5qdRkkLf1_fdPZK_9i80JDHNojW6 /view?usp=sharing

Tabel A. 6 Scrapping Script Produk Fashion Muslim Shopee

Kode scrapping produk fashion muslim Shopee	
U R L	https://drive.google.com/file/d/16LKWljdQUe_4zvgZmUj_ZKfFRNnyGTn2 /view?usp=sharing

Tabel A. 7 Scrapping Script Data URL UMKM Fashion Muslim Shopee

Kode scrapping data URL UMKM fashion muslim Shopee	
U R L	https://drive.google.com/file/d/1zVvZAoq96QSqI_Z_zBaw4WikriF7FLqS /view?usp=sharing

Tabel A. 8 Scrapping Script Informasi UMKM Fashion Muslim Shopee

Kode scrapping informasi UMKM fashion muslim shopee	
U R L	https://drive.google.com/file/d/1pT76x51K8OLHvoyP63RMrCD67AA0ZJQF /view?usp=sharing

3. Google My Business

Tabel A. 9 Scrapping Script Informasi UMKM Google My Business

Kode <i>scrapping</i> informasi UMKM google my business	
UR L	https://drive.google.com/file/d/1jb207BKDyOx27fsno2R1-5iIVbj2kxsb/view?usp=sharing

4. Instagram

Tabel A. 10 Scrapping Script Informasi Instagram

Kode <i>scrapping</i> informasi instagram dari nama UMKM	
U R L	https://drive.google.com/file/d/1ky1_kT0FxxTNFguUwOW_EY765Y2YT7k5/view?usp=sharing

B. Kode *Preprocess* data UMKM

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

1. Tokopedia

Tabel A. 11 Kode Preprocessing Informasi UMKM Fashion Muslim Tokopedia

Kode <i>preprocess</i> informasi UMKM fashion muslim di tokopedia	
U R L	https://drive.google.com/file/d/1kXgV2uHSnZYVy879_9BuE1bRhH2dj1TG/view?usp=sharing

2. Shopee

Tabel A. 12 Kode Preprocessing Informasi UMKM Fashion Muslim Tokopedia

Kode <i>preprocess</i> informasi UMKM fashion muslim di shopee	
U R L	<u>https://drive.google.com/file/d/1oisvLQpvdCNFvUYkRaG0GZRF1wQU_zJO/view?usp=sharing</u>

3. Instagram

Tabel A. 13 Kode Preprocessing Data Akun Instagram

Kode <i>preprocess</i> data akun di instagram	
U R L	<u>https://drive.google.com/file/d/1ky1_kT0FxxTNFguUwOW_EY765Y2YT7k5/view?usp=sharing</u>

Tabel A. 14 Kode Preprocessing Similarity Data Akun Instagram

Kode <i>preprocess</i> similarity data akun di instagram	
U R L	<u>https://drive.google.com/file/d/1zFT0yIBS7wuh2KUTMqpoTo5aU9va5bz/view?usp=sharing</u>

4. Google My Business

Tabel A. 15 Kode Preprocessing Informasi UMKM Google My Business

Kode <i>preprocess</i> informasi UMKM di Google My Business	
U R L	<u>https://drive.google.com/file/d/182YtE0NoyoIBvtTdxjMAYorRL4EVOB1/view?usp=sharing</u>

C. Kode Integrasi data UMKM

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

Tabel A. 16 Kode Integrasi Data UMKM Antara Tokopedia Dan Shopee

Kode integrasi data UMKM antara dua platform yaitu Tokopedia dan shopee		
U	https://drive.google.com/file/d/1U9MRgJVASS5GJJiN3fbtMBazAVTbR8I/view?usp=sharing	
R		
L		

Tabel A. 17 Kode Integrasi Data UMKM Antara Tokopedia, Shopee, dan Google My Business

Kode integrasi data UMKM antara tiga platform yaitu Google my business, Tokopedia dan shopee		
U	https://drive.google.com/file/d/1qwdhOsne0iCU7hI_eQSGDJOOEM6eLSGf/view?usp=sharing	
R		
L		

Halaman ini sengaja dikosongkan

LAMPIRAN B Hasil data

Pada bagian ini akan diberikan hasil data luaran *scrapping*, *preprocess* dan integrasi

A. Hasil data scrapping

1. Tokopedia

Tabel B. 1 Hasil Data Kategori Fashion Muslim Tokopedia

Hasil data kategori fashion muslim Tokopedia	
U R L	https://drive.google.com/file/d/15ORV23PvRYtpBavsou6XuABYCQC_qrR6/view?usp=sharing

Tabel B. 2 Hasil Data Produk Fashion Muslim Tokopedia

Hasil data produk fashion muslim Tokopedia	
U R L	https://drive.google.com/file/d/1nViohnzQg6K8kOKRHOKyagpRuKIUGC9n/view?usp=sharing

Tabel B. 3 Hasil Data URL UMKM Hasil Preprocess Kategori Fashion Muslim Tokopedia

Hasil data URL UMKM hasil <i>preprocess</i> kategori fashion muslim Tokopedia	
URL	https://drive.google.com/file/d/1223p48SXtzC-pTCZvUvNfLTsFwBr1I4P/view?usp=sharing

Tabel B. 4 Hasil Data Informasi UMKM Kategori Fashion Muslim Tokopedia

Hasil data informasi UMKM kategori fashion muslim Tokopedia	

UR L	https://drive.google.com/file/d/13AdtfGh_tJnLiKQ2YaC1r-MpHcSUszXy/view?usp=sharing
---------	---

2. Shopee

Tabel B. 5 Hasil Data Kategori Fashion Muslim Shopee

Hasil data kategori fashion muslim shopee	
UR L	https://drive.google.com/file/d/1TYaVBY9vm6OBwemKLLOk5JUOq-nGmgWl/view?usp=sharing

Tabel B. 6 Hasil data produk fashion muslim shopee

Hasil data produk fashion muslim shopee	
U R L	https://drive.google.com/file/d/1TGxy9jVQ9TRQf3yGP1EScIuL2VAFlW45/view?usp=sharing

Tabel B. 7 Hasil Data URL UMKM Fashion Muslim Shopee

Hasil data URL UMKM fashion muslim shopee	
U R L	https://drive.google.com/file/d/19JwzFc1yvDOfh_DdSYkGwGpdk3Jt7QyM/view?usp=sharing

Tabel B. 8 Hasil Data Informasi UMKM Fashion Muslim Shopee

Hasil data informasi UMKM fashion muslim shopee	
U R L	https://drive.google.com/file/d/1VSQ8LYE3hJCROmbKPV8b4afkkRRHCC34/view?usp=sharing

3. Instagram

Tabel B. 9 Hasil Data Instagram Shopee

Hasil data instagram shopee	
U R L	https://drive.google.com/file/d/1pkQ77r91eG_o7Qq84mzZBQ9vchDNCL4g/view?usp=sharing

Tabel B. 10 Hasil Data Instagram Tokopedia

Hasil data instagram tokopedia	
U R L	https://drive.google.com/file/d/1EsnPS8M0gUXK7I_jTYRVzwxPIBM0xmF7/view?usp=sharing

4. Google My Business

Tabel B. 11 Hasil Data Informasi UMKM Google My Business

Hasil data informasi UMKM google my business	
U R L	https://drive.google.com/file/d/1KrCy6sBlJT6HXqt1lZ_ZuBCIfoxtKcRr3/view?usp=sharing

B. Hasil data preprocess

Tabel B. 12 Data Informasi UMKM Fashion Muslim Tokopedia Hasil Preprocess

Data informasi UMKM fashion muslim Tokopedia hasil preprocess	
URL	https://drive.google.com/file/d/1b1IngdKfeEv9LBGM3_i7O4nQpGQY6n/view?usp=sharing

Tabel B. 13 Data Informasi UMKM Fashion Muslim Shopee Hasil Preprocess

Data informasi UMKM fashion muslim shopee hasil preprocess		
UR L	https://drive.google.com/file/d/1DqSyo9u2RP8Lv4OxiOB29U57-cKw75jG/view?usp=sharing	

Tabel B. 14 Data Informasi UMKM Fashion Muslim Google My Business Hasil Preprocess

Data informasi UMKM fashion muslim google my business hasil preprocess		
U R L	https://drive.google.com/file/d/1xwKloYnM2egBf7ZuDikjqak0d9wbqflU/view?usp=sharing	

Tabel B. 15 Data Akun Instagram dan Tokopedia Hasil Preprocess

Data akun instagram dan tokopedia hasil preprocess		
U R L	https://drive.google.com/file/d/1RUdgU6jZxnGDFRC9Sodg32ri06qPcmT0/view?usp=sharing	

Tabel B. 16 Data Akun Instagram dan Shopee Hasil Preprocess

Data akun instagram dan shopee hasil preprocess		
U R L	https://drive.google.com/file/d/10Y0A6s_z7bGQpTjiVt4HTe_1hkTIwsFL/view?usp=sharing	

C. Hasil data integrasi

Data UMKM hasil integrasi dibagi menjadi dua yaitu, data UMKM Tokopedia dengan shopee, dan google my business dengan Tokopedia dan shopee

Tabel B. 17 Data UMKM Fashion Muslim Hasil Integrasi Antara Tokopedia Dengan Shopee

Data UMKM fashion muslim hasil integrasi antara dua platform, Tokopedia dengan shopee	
U	https://drive.google.com/file/d/1DeUuS9E_OBHFE0ZjBCYwjrMScHB0dG95/view?usp=sharing
R	
L	

Tabel B. 18 Data UMKM Fashion Muslim Hasil Integrasi Antara Tiga Platform, Google My Business, Tokopedia Dan Shopee

Data UMKM fashion muslim hasil integrasi antara tiga platform, Google my business, tokopedia dan shopee	
U	https://drive.google.com/file/d/1wJd58vTJ0aCJ0zKr_6lCy2e5uEKZGp_W/view?usp=sharing
R	
L	

Halaman ini sengaja dikosongkan.

LAMPIRAN C Data Pendukung

a. Data kecamatan dan kelurahan

Tabel C. 1 Data Kecamatan dan Kelurahan

Data kecamatan dan kelurahan	
U R L	https://drive.google.com/file/d/1H-a2TJLGJvz3KawPETWY4D242oC2nFD8/view?usp=sharing

b. Data kategori

Tabel C. 2 Data Kategori Google Business

Data kategori google business	
U R L	https://drive.google.com/file/d/1VEB6PuPJZSQ24brFVHSNNwlaXZPPgKR8/view?usp=sharing

c. Data kategori relevan fashion muslim

Tabel C. 3 Data Preprocessing Pemilihan Kategori

Data preprocessing pemilihan kategori	
U R L	https://drive.google.com/file/d/1JgXPVLx6_x6R9lsJJs3JoTb0mPGwnHz3/view?usp=sharing

Halaman ini sengaja dikosongkan.

LAMPIRAN D Tabel Chi Kuadrat

Tabel D. 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
12	10.34100	17.27501	19.67514	24.72497
13	11.34032	18.54935	21.02607	26.21697
14	12.33976	19.81193	22.36203	27.68825
15	13.33927	21.06414	23.68479	29.14124

BIODATA PENULIS



Penulis lahir di Surabaya pada tanggal 25 Mei 1998 dan merupakan anak kedua dari dua bersaudara pasangan Bapak Nasrudin dan Ibu Sugiarti Arifiani. Penulis menempuh Pendidikan formal di SD Negeri Rungkut Menanggal II Surabaya, SMP Negeri 35 Surabaya dan SMA Negeri 20 Surabaya.

Pada tahun 2016 penulis melanjutkan Pendidikan jenjang sarjana dengan jalur PKM di

Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas(FTIK)-Institut Teknologi Sepuluh Nopember(ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 05211640000134. Selama masa perkuliahan penulis aktif mengikuti organisasi mahasiswa, seperti Kajian Islam Sistem Informasi, Jamaah Masjid Manarul Ilmi ITS. Penulis juga aktif dalam kegiatan kepanitiaan, kegiatan keilmiahian dan kegiatan sosial, seperti ISE, FTIf Festival, Lomba *Startup*, konferensi internasional, kegiatan *entrepreneurship*.

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