

BAB III

METODOLOGI PENELITIAN

3.1 Lokasi Penelitian

Lokasi penelitian ini adalah Taman Nasional Bromo Tengger Semeru (TNBTS) yang berada di Jawa Timur. Secara administratif, TNBTS terletak di Kabupaten Malang, Kabupaten Pasuruan, Kabupaten Probolinggo, dan Kabupaten Lumajang, Provinsi Jawa Timur. TNBTS terletak pada puncak ketinggian 750 hingga 3.676 mdpl, yang merupakan puncak Gunung Semeru, puncak tertinggi di Pulau Jawa.

Berada di area pegunungan dan memiliki banyak potensi alam dan budaya menjadikan area ini sebagai salah satu kawasan konservasi yang harus dilindungi. Hal ini juga yang membuat wisatawan tertarik untuk berkunjung ke TNBTS. TNBTS dikelola oleh Balai Besar Taman Nasional Bromo Tengger Semeru (di bawah Kementerian Lingkungan Hidup dan Kehutanan) dan Provinsi Jawa Timur.

3.2 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode survei. Pendekatan kuantitatif digunakan karena penelitian ini dilakukan dengan analisis data yang bersifat statistik. Metode survei adalah metode yang menggunakan kuesioner sebagai instrumen utama untuk mengumpulkan data (Siyoto & Sodik, 2015). Penelitian dilakukan secara *virtual* dengan menyebar kuesioner melalui *google form* yang kemudian disebar melalui media sosial kepada responden.

Teknik analisis yang digunakan adalah analisis *cluster*. Analisis *Cluster* merupakan analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya (Eddie et.al., 2013). Analisis *cluster* digunakan karena penelitian ini hanya mempunyai satu variabel, yaitu tipologi yang berisi beberapa sub variabel di dalamnya.

3.3 Populasi dan Sampel

Menurut Sugiyono (dalam Siyoto & Sodik, 2015), populasi merupakan wilayah generalisasi yang terdiri dari obyek atau subyek yang memiliki kuantitas

dan karakteristik tertentu yang ditentukan peneliti untuk dipelajari dan kemudian ditarik kesimpulan. Populasi dari penelitian ini merupakan wisatawan nusantara Taman Nasional Bromo Tengger Semeru pada rentang tahun 2020-2021 yang mengunggah foto atau videonya di media sosial. Rentang tahun 2020 hingga 2021 digunakan oleh penulis karena pada tahun 2020 hingga 2021 merupakan masa di mana wisata alam menjadi salah satu tren wisata prioritas di Indonesia (kemenparekraf.go.id).

Populasi penelitian ini berdasarkan jumlah unggahan yang menggunakan tagar yang berkaitan dengan Taman Nasional Bromo Tengger Semeru. Berikut ini merupakan tabel jumlah unggahan yang menggunakan tagar yang berkaitan dengan Taman Nasional Bromo Tengger Semeru. Berdasarkan tabel di bawah, terdapat 1.683.000+ unggahan yang berkaitan dengan TNBTS. Jumlah unggahan pada sosial media tersebut menunjukkan eksistensi TNBTS sebagai salah satu tujuan wisata yang populer.

Tabel 3.1 Tagar TNBTS di Instagram

Tagar	Jumlah Unggahan
#bromotenggersemeru	708.000+
#bromomountain	970.000+
#wonderfulbromo	5.000+
Total	1.683.000+

Sumber: Hasil Olah Data Penulis (2022)

Sampel merupakan sebagian dari jumlah dan karakteristik yang dimiliki oleh populasi (Siyoto & Sodik, 2015). Sampel yang akan digunakan harus benar-benar dapat mewakili populasi yang diteliti (Sugiyono, 2013). Teknik sampling yang digunakan, yaitu *purposive sampling*. Teknik *purposive sampling* merupakan suatu teknik penentuan sampel yang memiliki kriteria tertentu (Siyoto & Sodik, 2015). Berikut merupakan tabel jumlah kunjungan wisatawan nusantara TNBTS pada rentang tahun 2020-2021. Tabel ini digunakan untuk lebih memperkecil jangkauan responden.

Tabel 3.2 Jumlah Kunjungan Wisatawan Nusantara TNBTS

Tahun	Jumlah Kunjungan
2020	196.268
2021	138.695
Jumlah	334.963

Sumber: KLHK (2021); KLHK (2022)

Kriteria pemilihan sampel pada penelitian ini adalah wisatawan nusantara yang pernah berkunjung ke TNBTS dan mengunggah dokumentasinya saat berlibur di media sosial serta berusia minimal 17 tahun, karena pada usia tersebut responden dianggap sudah dapat memahami dan menjawab pertanyaan pada kuesioner. Alasan mengenai kriteria wisatawan yang berkunjung ke TNBTS dan mengunggahnya di media sosial adalah sebagai salah satu penentuan dalam penilaian variabel motivasi prestise dalam penelitian ini.

Penentuan jumlah sampel dalam penelitian ini penulis menggunakan rumus Slovin. Rumus Slovin digunakan karena jumlah populasi penelitian ini cukup besar, sehingga diperlukan jumlah sampel yang tepat dan dapat mewakili keseluruhan populasi. Rumus Slovin menggunakan 10-20% sampel dari total keseluruhan populasi. Berikut merupakan penjelasan rumus Slovin:

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n = jumlah sampel minimal

N = jumlah populasi

e = *margin of error* (batas toleransi kesalahan)

Rumus Slovin memiliki ketentuan sebagai berikut:

- a. Nilai $e = 0,1$ (10%) untuk populasi dalam jumlah besar
- b. Nilai $e = 0,2$ (20%) untuk populasi dalam jumlah kecil

Jumlah populasi dalam penelitian ini adalah sebanyak 334.963 orang. Jumlah tersebut diperoleh dari jumlah total wisatawan nusantara TNBTS selama periode

tahun 2020 – 2021. Sehingga persentase *margin of error* (e) yang digunakan adalah 10% (0,1) dan hasil perhitungannya akan dibulatkan untuk mencapai kesesuaian. Berikut adalah perhitungan pengambilan sampel penelitian.

$$n = \frac{334963}{1 + 334963(0.1)^2}$$

$$n = \frac{334963}{3350.63}$$

$$n = 99.97 ; \text{ dibulatkan menjadi } 100 \text{ sampel}$$

Berdasarkan perhitungan di atas, didapatkan nilai sampel minimal yang menjadi responden dalam penelitian ini dibulatkan menjadi sebanyak 100 sampel. Penyebaran sampel ini dilakukan dengan mengirimkan tautan kuesioner *google form* kepada wisatawan nusantara yang pernah berkunjung ke Taman Nasional Bromo Tengger Semeru melalui media sosial *Instagram*.

3.4 Sumber Data

Data merupakan informasi yang dibutuhkan oleh peneliti untuk memecahkan masalah atau menjawab pertanyaan penelitian (Siyoto & Sodik, 2015). Data yang digunakan penulis dalam melakukan penelitian ini ada dua, yakni data primer dan data sekunder. Data primer merupakan sumber data yang diperoleh secara langsung, ke lokasi atau objek penelitian. Dalam penelitian ini data primer diperoleh dari hasil kuesioner yang disebarakan secara daring melalui sosial media, yaitu *Instagram*. Sedangkan data sekunder merupakan data yang didapat oleh peneliti dari sumber yang sudah ada, seperti *ebook*, artikel jurnal, artikel berita, situs *web* dan sosial media resmi Balai Besar Taman Nasional Bromo Tengger Semeru serta sosial media resmi Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia untuk mendapatkan informasi terkait jumlah pengunjung, batas administratif dan informasi pendukung lainnya.

3.5 Operasional Variabel Penelitian

Variabel penelitian merupakan segala sesuatu dalam bentuk apapun yang ditetapkan oleh peneliti untuk dipelajari sehingga diperoleh informasi dan kemudian ditarik kesimpulannya (Sugiyono, 2013). Pada penelitian ini terdapat empat variabel, yaitu Pilihan Produk, Aktivitas Vandalisme, Motivasi Prestise, dan

Pengetahuan Wisatawan. Secara rinci, operasional variabel dalam penelitian ini terdapat pada tabel berikut:

Tabel 3.3 Operasional Variabel

Variabel	Aspek	Indikator	Skala	Kode
Tipologi				
Pilihan Produk	Paket wisata	Jenis paket wisata yang dipilih	<i>Likert</i>	PP1
	Akomodasi	Jenis akomodasi yang dipilih		PP2
	<i>Nature</i>	Pilihan destinasi wisata yang masih alami		PP3
	Kelompok wisata	Jenis kelompok perjalanan wisata yang dipilih		PP4
Aktivitas Vandalisme	<i>Delayed of damage</i>	Kecenderungan untuk mengambil flora/fauna yang dilindungi	<i>Likert</i>	A1
		Kecenderungan untuk mengambil spesimen yang terdapat di kawasan konservasi		A2
Motivasi Prestise	<i>Uniqueness</i>	Keinginan berwisata untuk meningkatkan kepercayaan diri dengan mengunjungi tempat yang unik	<i>Likert</i>	MP1
	<i>Self-esteem</i>	Berwisata untuk meningkatkan kepercayaan diri dengan mengunjungi tempat yang populer		MP2

Variabel	Aspek	Indikator	Skala	Kode
	<i>Recognition</i>	Berwisata untuk mendapatkan pengakuan dari orang lain		MP3
Pengetahuan Wisatawan	<i>Awareness of Consequence</i>	Kekhawatiran akan dampak aktivitas pariwisata terhadap ekosistem kawasan konservasi	<i>Likert</i>	TK1
		Kekhawatiran akan dampak aktivitas pariwisata terhadap lingkungan alam kawasan konservasi		TK2
	<i>Knowledge statements</i>	Pengetahuan wisatawan mengenai cara untuk mendukung konservasi		TK3

Sumber: Hasil Olah Data Penulis (2022)

3.6 Instrumen Penelitian

Instrumen penelitian digunakan untuk mengukur nilai variabel yang diteliti. Instrumen penelitian adalah alat yang digunakan untuk mengukur fenomena alam atau sosial yang diamati (Sugiyono, 2013). Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode angket atau kuesioner, instrumennya berupa kuesioner. Kuesioner penelitian ini berisi 12 pernyataan indikator dengan menggunakan skala likert sebagai jawabannya. Skala likert adalah acuan untuk mengukur sikap, pendapat, dan persepsi seseorang, atau sekelompok orang tentang fenomena sosial (Sugiyono, 2013).

Selain pernyataan indikator, terdapat beberapa bagian lain yang harus diisi oleh responden, yaitu karakteristik responden, seperti jenis kelamin, usia, pendidikan terakhir, domisili, pekerjaan, penghasilan, tujuan kunjungan, frekuensi kunjungan, pola kunjungan, serta sumber informasi mengenai destinasi. Kemudian, bagian kedua berisi pernyataan-pernyataan indikator mengenai pilihan produk, aktivitas vandalisme, motivasi prestise, dan pengetahuan wisatawan yang

menggunakan pengukuran skala likert. Skala likert yang digunakan dapat dilihat pada **Tabel 3.4**.

Tabel 3.4 Skala Likert

Skor	Skala Jawaban
5	Sangat Setuju
4	Setuju
3	Cukup
2	Tidak Setuju
1	Sangat Tidak Setuju

Sumber: Sugiyono (2013)

3.7 Teknik Pengumpulan Data

Dalam penelitian ini diawali dengan studi pustaka, peneliti mengkaji beberapa jurnal yang berkaitan dengan topik penelitian, yaitu analisis tipologi wisatawan. Kemudian, peneliti menentukan lokasi objek penelitian, yaitu Taman Nasional Bromo Tengger Semeru dan populasi serta sampel penelitian ini adalah wisatawan yang pernah mengunjungi TNBTS. Setelah populasi dan sampel ditentukan, kemudian menyebarkan kuesioner secara daring, melalui media sosial Instagram dengan fitur tagar serta pesan (*Direct Message*).

3.8 Teknik Pengujian Data

3.8.1 Uji Validitas

Uji validitas digunakan untuk mengetahui apakah alat ukur tersebut dapat dengan tepat mengukur apa yang hendak diukur (Sugiyono, 2007). Dalam uji validitas setiap pernyataan diukur dengan menghubungkan total dari setiap pernyataan dengan total keseluruhan tanggapan setiap pernyataan (Darma, 2021).

Kriteria uji validitas adalah sebagai berikut:

- a. Jika nilai r hitung $>$ r tabel, maka data dinyatakan valid
- b. Jika nilai r hitung $<$ r tabel, maka data dinyatakan tidak valid

Nilai signifikansi:

- a. Jika nilai signifikansi $<$ $0.05 =$ Valid

b. Jika nilai signifikansi >0.05 = Tidak Valid

Uji validitas penelitian ini menggunakan aplikasi IBM SPSS V.26 dengan teknik Pearson. Jumlah responden yang digunakan dalam uji validitas ini sebanyak 30 responden. Berdasarkan hasil pengamatan r tabel, jika $N=30$, maka skor r tabel = 0,361. Tabel di bawah ini merupakan hasil uji validitas untuk variabel pilihan produk, aktivitas vandalisme, motivasi prestise, dan pengetahuan/akses konservasi.

Tabel 3.5 Hasil Uji Validitas

Variabel	Kode	r hitung	r tabel	Keterangan
Pilihan Produk	PP1	0,830	0,361	Valid
	PP2	0,609	0,361	Valid
	PP3	0,609	0,361	Valid
	PP4	0,837	0,361	Valid
Aktivitas Vandalisme	A1	0,709	0,361	Valid
	A2	0,662	0,361	Valid
Motivasi Prestise	MP1	0,896	0,361	Valid
	MP2	0,915	0,361	Valid
	MP3	0,867	0,361	Valid
Pengetahuan Wisatawan	TK1	0,919	0,361	Valid
	TK2	0,963	0,361	Valid
	TK3	0,795	0,361	Valid

Sumber: Hasil Olah Data Penulis (2022)

Berdasarkan tabel di atas, dinyatakan bahwa setiap item pernyataan terbukti valid, sebab nilai r hitung lebih besar daripada nilai r tabel. Sehingga jumlah item pada instrumen penelitian ini tidak terdapat pengurangan dan sah untuk digunakan dalam penelitian.

3.9 Teknik Analisis Data

Teknik analisis data penelitian ini adalah analisis kluster (*cluster analysis*). Analisis kluster adalah analisis multivariat yang digunakan untuk mengelompokkan objek menjadi suatu kelompok dengan kriteria tertentu yang sama (Idrus, 2013). Analisis kluster mengelompokkan objek yang memiliki kemiripan dengan objek lain ke dalam suatu kluster. Hasil kluster suatu objek harus memiliki homogenitas

yang tinggi dalam kluster yang sama dan memiliki perbedaan yang tinggi terhadap objek pada kluster yang lain (Wijaya & Budiman, 2013).

Analisis kluster memiliki dua metode di dalamnya, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. Salah satu metode kluster non-hirarki adalah *K-Means clustering*. *K-Means Clustering* adalah metode yang mengelompokkan objek ke dalam k kluster berdasarkan jarak centroid terdekat. Metode ini merupakan metode yang paling sering digunakan. Proses analisis penelitian ini menggunakan aplikasi IBM SPSS Statistics 26 untuk memudahkan penulis dalam mengolah data. Pengolahan data penelitian ini terbagi menjadi beberapa tahap, yaitu:

1. *Pre-processing Data*

Pre-processing data merupakan tahapan awal dalam proses data *mining*. Tahap ini terbagi menjadi dua tahapan di dalamnya, yaitu:

- a. Seleksi Data

Seleksi data bertujuan untuk menyeleksi data yang digunakan karena tidak semua yang akan digunakan, hanya data yang sesuai untuk dianalisa yang akan diambil dari *database*. *Database* yang dimaksud dalam penelitian ini adalah data yang berisi jawaban responden pada kuesioner (*Google Form*) yang terbentuk dalam format *file Microsoft Excel*. Berikut adalah gambar yang menunjukkan *database* penelitian ini.

Pada **gambar 3.1** menunjukkan bahwa data masih mentah, data masih berupa hasil langsung dari responden yang mengisi kuesioner, pada tahap seleksi data ini memilih data yang tidak digunakan, hal ini dilakukan untuk memperkecil ruang lingkup penelitian. Kuesioner dalam penelitian ini berisi 22 pertanyaan, 10 pertanyaan profil responden dan 12 lainnya mengenai variabel penelitian. Namun, data yang akan diolah dengan menggunakan metode *K-Means Clustering* hanya 12 pertanyaan terkait variabel penelitian, yaitu 4 pertanyaan mengenai variabel pilihan produk, 2 pertanyaan mengenai aktivitas vandalisme, 3 pertanyaan mengenai motivasi prestise, serta 3 pertanyaan mengenai pengetahuan wisatawan.

Timestamp	Email Address	Nama	Jenis Kelamin	Usia	Pendidikan Terakhir	Domisili	Pekerjaan	Penghasilan Per Bulan
3/29/2022 11:44:38	darapatloka74@gmail.com	Dara	Wanita	26 - 35 tahun	Perguruan Tinggi	Semarang	Wiraswasta	Rp1.500.001-Rp2.500.000
3/29/2022 11:55:59	yukowie@gmail.com	Yeka	Pria	36 - 45 tahun	Perguruan Tinggi	Surabaya	Wiraswasta	>Rp3.500.000
3/29/2022 13:45:31	fikasmarini@gmail.com	Chira	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	DKI Jakarta	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 20:59:28	agbeldhena83@gmail.com	Agbel dhena	Wanita	17 - 25 tahun	Perguruan Tinggi	Sragen	Pegawai Swasta	Rp2.500.001-Rp3.500.000
3/29/2022 21:51:42	your.ours0@gmail.com	ia	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	Jakarta Timur	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 22:00:47	putriayukharisma123@gmail.com	Putri Ayu	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	Tangerang	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 22:25:29	piecewines@gmail.com	China	Wanita	17 - 25 tahun	Perguruan Tinggi	Tangerang	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 22:33:14	albertadonette@gmail.com	Alberta	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	Jakarta	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 22:44:06	vinayawinandaa@gmail.com	vinaya	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	tangerang	Pelajar/Mahasiswa	<Rp1.500.000
3/29/2022 23:35:21	andinakara@gmail.com	kara	Wanita	17 - 25 tahun	Perguruan Tinggi	Malang	Mahasiswa	>Rp3.500.000
3/29/2022 12:43:28	leoridho260701@gmail.com	Leo ridho	Pria	17 - 25 tahun	SMA/SMK/ sederajat	Jawatehng	Pelajar/Mahasiswa	<Rp1.500.000
3/31/2022 9:03:07	lilik.darmasto@gmail.com	LILIK DARMASTO	Pria	26 - 35 tahun	Perguruan Tinggi	SOLO	Pegawai Swasta	>Rp3.500.000
3/31/2022 12:35:40	arohmasukmapmd4@gmail.com	Arrohma Sukma	Wanita	17 - 25 tahun	Perguruan Tinggi	Yogyakarta	Pegawai Swasta	>Rp3.500.000
3/31/2022 13:23:53	trio.candra87@gmail.com	Trio	Pria	17 - 25 tahun	SMA/SMK/ sederajat	Malang	Pegawai Swasta	<Rp1.500.000
3/31/2022 13:27:52	fhdabai@gmail.com	Farah Diba	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	Kediri	Pelajar/Mahasiswa	<Rp1.500.000
3/31/2022 13:54:55	maskhud22@gmail.com	Maskhud Yunus	Pria	26 - 35 tahun	Perguruan Tinggi	Jawa Timur	Pegawai Swasta	>Rp3.500.000
3/31/2022 13:57:55	ricky7337@gmail.com	Ricky	Pria	17 - 25 tahun	Perguruan Tinggi	Bogor	Pelajar/Mahasiswa	<Rp1.500.000
3/31/2022 14:11:06	nurulkhani94@gmail.com	Khanif	Pria	26 - 35 tahun	Perguruan Tinggi	Tangerang	Pegawai Swasta	>Rp3.500.000
3/31/2022 15:03:33	reskihasnani@gmail.com	Reski Amalia	Wanita	17 - 25 tahun	SMA/SMK/ sederajat	Binrang	Pelajar/Mahasiswa	<Rp1.500.000

Gambar 3.1 Database

Sumber: Hasil Olah Data Penulis, 2022

Tabel berikut merupakan hasil data yang telah diseleksi, hasil data berisi 12 pertanyaan indikator yang akan diolah menggunakan metode *K-means Clustering*.

Tabel 3.6 Seleksi Data

Variabel	Aspek	Kode
Pilihan Produk	Paket Wisata	PP1
	Akomodasi	PP2
	Nature	PP3
	Kelompok Wisata	PP4
Aktivitas Vandalisme	Delayed of damage	A1
		A2
Motivasi Prestise	Uniqueness	MP1
	Self-esteem	MP2
	Recognition	MP3
Pengetahuan Wisatawan	Awareness of consequence	TK1
		TK2
	Problem awareness and Knowledge statements	TK3

Sumber: Hasil Olah Data Penulis (2022)

Gita Kiranti, 2022

ANALISIS TIPOLOGI WISATAWAN DI KAWASAN KONSERVASI TAMAN NASIONAL BROMO TENGER SEMERU

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

b. Reduksi Data

Reduksi data merupakan proses pengurangan jumlah data yang tidak diperlukan selama proses analisis. Tujuan proses ini adalah untuk memperoleh data yang lebih relevan. Sehingga, data yang telah direduksi akan memberikan gambaran yang lebih jelas dan mempermudah dalam proses pengolahan data menggunakan metode *K-Means Clustering*. Pada tahap ini data yang dikurangi yaitu 10 pertanyaan mengenai profil responden. Berikut merupakan gambar hasil reduksi data, di mana data mengenai profil responden sudah dikurangi.

Tourist ID	PP_1	PP_2	PP_3	PP_4	A_1	A_2	A_3	A_4	A_5	MP_1	MP_2	MP_3	TK_1	TK_2	TK_3
1	4	3	3	3	2	2	3	5	3	4	4	4	3	3	3
2	5	3	5	5	2	5	3	5	5	5	5	3	1	1	1
3	4	2	4	4	1	2	2	5	4	3	3	3	1	1	3
4	2	3	3	1	1	1	3	4	4	3	3	3	4	4	4
5	4	4	4	4	2	4	4	4	4	4	4	4	4	4	4
6	4	4	3	4	2	2	3	4	4	4	3	2	2	3	4
7	4	4	3	4	2	2	2	4	4	3	3	3	3	3	4
8	5	4	5	4	1	1	4	5	4	4	5	4	2	2	5
9	5	3	3	3	2	3	1	1	3	2	1	1	2	2	5
10	3	1	4	1	1	2	1	4	3	2	2	2	5	5	5
11	3	3	4	1	1	2	4	3	2	3	4	3	4	4	4
12	3	2	4	1	1	1	2	1	2	5	5	4	5	5	5
13	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
14	5	2	3	4	1	1	2	5	3	4	3	3	3	3	4
15	4	3	1	3	2	5	2	5	2	5	5	3	4	5	5
16	4	3	5	2	1	1	1	2	2	3	3	3	5	5	5
17	2	2	1	3	1	1	2	3	2	3	2	1	4	4	5
18	4	3	4	2	2	2	2	2	3	3	3	4	4	4	4
19	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
20	5	4	4	5	1	1	5	5	5	5	5	5	5	5	5
21	4	5	2	1	1	3	2	5	4	2	2	2	3	4	5
22	5	4	5	4	3	3	4	5	3	5	5	4	4	4	5

Gambar 3.2 Hasil Reduksi Data

Sumber: Hasil Olah Data Penulis (2022)

c. Pembersihan Data (*Data Cleaning*)

Data yang telah direduksi kemudian masuk pada tahap pembersihan data. Pembersihan dilakukan untuk membuang data yang tidak layak untuk memasuki proses data mining, seperti data yang mengandung *outlier* (tidak sesuai dengan penelitian), ganda, missing values atau tidak lengkap. Maka, dilakukan pengecekan ulang pada data hasil tanggapan responden.

c. Transformasi Data

Pada tahap ini dilakukan perubahan data yang memiliki tipe data yang awalnya tidak bisa diolah secara matematis menjadi data yang bisa diolah. Tujuan dari transformasi data untuk menghindari data yang rusak dan tidak valid. Pada penelitian ini, proses transformasi data yang dilakukan adalah

menghitung rata-rata dari masing-masing variabel penelitian. Gambar di bawah ini merupakan hasil transformasi data, di mana pada bagian kolom yang berwarna merupakan hasil perhitungan rata-rata setiap variabel penelitian, yaitu Pilihan Produk, Aktivitas, Motivasi Prestise, dan Pengetahuan Wisatawan.

	PP_1	PP_2	PP_3	PP_4	Mean	A_1	A_2	A_3	A_4	A_5	Mean	MP_1	MP_2	MP_3	Mean	TK_1	TK_2	TK_3	Mean
2	4	3	3	3	3.25	2	1	3	5	3	3.00	4	4	4	4.00	3	3	3	3.00
3	5	3	5	5	4.50	2	5	3	5	5	4.00	5	5	3	4.33	1	1	1	1.00
4	4	2	4	4	3.50	1	2	2	5	4	2.80	3	3	3	3.00	1	1	3	1.67
5	2	3	3	1	2.25	1	1	3	4	4	2.60	3	3	3	3.00	4	4	4	4.00
6	4	4	4	4	4.00	2	4	4	4	4	3.60	4	4	4	4.00	4	4	4	4.00
7	4	4	3	4	3.75	2	2	3	4	4	3.00	4	3	2	3.00	2	3	4	3.00
8	4	4	3	4	3.75	2	2	2	4	4	2.80	3	3	3	3.00	3	3	4	3.33
9	5	4	5	4	4.50	1	1	4	5	4	3.00	4	5	4	4.33	2	2	5	3.00
10	5	3	3	3	3.50	2	3	1	1	3	2.00	2	1	1	1.33	2	2	5	3.00
11	3	1	4	1	2.25	1	2	1	4	3	2.20	2	2	2	2.00	5	5	5	5.00
12	3	3	4	1	2.75	1	3	4	3	2	2.40	3	4	3	3.33	4	4	4	4.00
13	3	2	4	1	2.50	1	1	2	1	2	1.40	5	5	4	4.67	5	5	5	5.00
14	5	5	5	5	5.00	5	5	5	5	5	5.00	5	5	5	5.00	5	5	5	5.00
15	5	2	3	4	3.50	1	1	2	5	3	2.40	4	3	3	3.33	3	3	4	3.33
16	4	3	1	3	2.75	2	5	2	5	2	3.20	5	5	3	4.33	4	5	5	4.67
17	4	3	5	2	3.50	1	1	1	2	2	1.40	3	3	3	3.00	5	5	5	5.00
18	2	2	1	3	2.00	1	1	2	3	2	1.80	3	2	1	2.00	4	4	5	4.33
19	4	3	4	2	3.25	2	2	2	2	2	2.00	3	3	3	3.00	4	4	4	4.00
20	4	4	4	4	4.00	4	4	4	4	4	4.00	4	4	4	4.00	4	4	4	4.00
21	5	4	4	5	4.50	1	1	3	5	5	3.40	5	5	5	5.00	5	5	5	5.00
22	4	5	2	1	3.00	1	3	2	5	4	3.00	2	2	2	2.00	3	4	5	4.00
23	4	5	2	1	3.00	1	3	2	5	4	3.00	2	2	2	2.00	3	4	5	4.00

Gambar 3.3 Transformasi Data

Sumber: Hasil Olah Data Penulis (2022)

2. Proses Perhitungan K-Means Clustering

Tahap ini adalah tahapan yang menerapkan metode dari data mining untuk mengolah data yang ada. Metode yang digunakan adalah *K-Means Clustering*, proses kerja metode ini adalah dengan menghitung jarak setiap data ke pusat *cluster* antara objek ke *centroid* dengan perhitungan jarak euclidean dengan rumus pada persamaan. Namun, dalam penelitian ini menggunakan aplikasi IBM SPSS Statistics 26, di mana perhitungan *cluster* dilakukan secara sistematis. *Output* yang didapat adalah beberapa kelompok wisatawan yang berkunjung ke Taman Nasional Bromo Tengger Semeru.

3. Interpretasi/Evaluasi

Interpretasi/Evaluasi adalah tahap untuk melakukan evaluasi hasil dari penelitian setelah melakukan proses pengolahan data. Evaluasi dilakukan untuk menerjemahkan hasil metode clustering dalam format yang lebih mudah dipahami. Proses ini dilakukan dengan menginterpretasikan hasil dengan teori penelitian pada setiap cluster.