



Bersinergi dengan Data,
Majukan Pariwisata Nusantara

LAPORAN HASIL

Nomor Buku 10.LH



Riset 1

Kajian Pemanfaatan
Mobile Positioning Data (MPD)
dalam Menunjang Statistik Pariwisata

Praktik Kerja Lapangan D-IV
Politeknik Statistika STIS
T.A. 2022/2023

Penanggung Jawab Umum	: Dr. Erni Tri Astuti, M.Math
Penanggung Jawab Akademik	: Prof. Setia Pramana S.Si, M.Sc, Ph.D
Penanggung Jawab Keuangan	: Dr. Hardius Usman, M.Si
Penanggung Jawab Kemahasiswaan	: Yunarso Anang Sulistiadi, M.Eng.,Ph.D.
Penanggung Jawab Adm. Akademik	: Nurseto Wisnumurti, M.Stat
Penanggung Jawab Adm. Umum	: Bambang Nurcahyo, SE, MM
Ketua Penyelenggara	: Dr. Eng. Arie Wahyu Wijayanto, SST, MT
Koordinator	: Dr. Siti Muchlisoh, M.Si
Sekretaris	: Wahyuni Andriana Sofa, SST, MIDECA
Bendahara	: Rina Hardiyanti, SST.
Narasumber	: Dr. Azka Ubaidillah, SST.,M.Si Ibnu Santoso, SST.,MT
	Tim <i>Mobile Positioning Data</i> (MPD) BPS
Koordinator Pembimbing Riset	: Dr.Eng. Lya Hulliyyatus Suadaa, SST.,M.T
Tim Dosen	: Rahadi Jalu Yoga Utama, S.Tr.Stat. Budi Yuniarto, SST., M.Si Dr. Timbang Sirait, M.Si Lutfi Rahmatuti Maghfiroh, SST., MT Yuliagnis Transver Wijaya, S.ST, M.Sc
	Rani Nooraeni, SST, M.Stat Dr. Drs. Waris Marsisno M.Stat Robert Kurniawan, SST., M.Si Ibnu Santoso, SST., MT Dewi Kusuma Hartati, SST Juniarita Ratnasari, S.Tr.Stat.
	Dr. Bony Parulian Josaphat, S.Si.,M.Si Dr. Achmad Syahrul Choir, SST.,M.Si



Tim Mahasiswa :

BPH

Firhand Yusuf Kamal
Rofa Raudhatul Jannah
Lasta Sabila Krisyolin

Khoirunisa Maula Izzaty

Mufqi Ananda Rahmadi

Rafaela Suryadiva Indira Maharani

Rafif Hasabi

Syafni Dwi Darmawan

Rifqi Ramadhan

Nur Ainun Daulay

Anak Agung Gede Rai Bhaskara
Darmawan Pemayun

Desy Ahyu Fitria

Fatya Ratnadilla

Isnaini Rahma Dewi

Mesya Anggita Robiul Awaliah

Nandya Rezky Utami

Nur Hidayah Apriliani

Raditya Amanta Putra

Ramadhanti Rosyada

Taufiq Agung Kurniawan

Seksi Analisis

Muhammad Farhan
Diva Alfiah Hakim
I Made Satria Ambara Putra
Widhelia Echa Pramesthy
Alya Rudita Priyandini
Andriansyah Ichwan
Arnanda Syahputra
Fika Fuza Syahdana
Hanny Nur Rahma
Navika Artiari
Nisa Ayu Larasati
Rayhan Zaki Abrar
Zahra Ayu Lestari
Giani Jovita Jane
Widia Astuti
Aulia Salsabilla
Berliana Sugiarti Putri
Figri Al Rasjid Abdullah
Kamal Maulana Alfi
Karina Cindy Rahmanto

Seksi Metodologi

Ladisa Busaina
Khasatun Prasasti Saputri
Calvin Timothy Panjaitan
Salsabila Shafa Putri Rahadian
Berliana Lolita Siahaan
Charisma Noorraya Kusuma
Farkhan Ariyasa



Firrar Ayu Hastungkara Sudrajat
Jauharotul Maslakhah
Mohamad Hozinul Asror
M. Ziko Azizi
Mutiara Indryan Sari
Retno Palupi
Septian Tegar Yuristiawan
Shashella Zelicha Ramadhanty
Vivi Dewita Sari
Erfangga Bayu Kurniawan
Isna Syafira Khalif Ilma
Ade Rifqy Setyawan
Ages Marcello
Avelia Deavy Putri Primadani
Brillian Jeremia Sianturi
Deanty Fatihatul Magfirah
Desy Shintawati
Fitryanto Nur Hidayah
Muhammad Rafii Al Muflih
Muhammad Mishbah Nur Ihsan
M. Taufiqqurrahman
Muhammad Zaki Ramdhani
Rizky Wahyuda Manik
Rizquna Nurul Fatihah
Sukma Andini
Theresia Romatua

Seksi Kuesioner
Intan Trihandini Alawiyah
Masna Novita Rahmani
Siti Larasati
Afysya Diadara
Adelina Rahmawati
Bayu Rayhan Anassai
Ruri Tri Wahyuni
Ii Desika
Ahmad Faqih Pratama
Roby Awaludin Fajar
Alisha Azri Khairani
Putri Muthi'ah Dini
Ervan N R
Ahmad Tegar Prananggalih
Umu Arifatul Syifa
Fathna Amira
Nur Eka Diana
Restu Satrio Pinanggih
Norvan Bagus Ramadhan
Atina Asyfa Amania
Fitri Handayani
Rarisza Nabila
Hana Raihanatul Jannah
Laila Arrohmah
Fika Sabilha Haq
Tiara Ayu Sihurina



Saifullah Fatah
Nur Retno Fitriyyah
Natanael Agave Sitorus
Muhammad Afnan Falieh
Annisa Syifa Faulhaq
Hatmi Yutika
Jiddan Najdi Putra Adinata
Lutfiah Kumala Dewi

Seksi Pengolahan

Gilang Wahyu Prasetyo
Ahmad Sovi Hidayat
Faiq Rosadi Arridho
Wa Ode Raksni Dumpatua
Akhir Mubarok
Atikah Salsabila
Syifannisa Rahma Aulia
Theresia Intan Jessica

Rahma Fajri Ramdhani
Sinatrya Dwi Purnatadya
Haidar Hilmy Ahmady
Muhammad Rayyan Maulana
Thomson
Vivi Adelia Ramadian M
Zhafara Alfa Abdillah
Amanda Christina Saragih
Anisa Hardiani Putri
Lintang Wigati
Satya Indra Kusworo
Komang Niko Romano Prodi
Zhasa Vashellya
Pandya Almastiana Alifka
Khairiyah Nurindah Yudha Maharani
Michael Jonas.S



PRAKATA

Puji syukur kehadirat Tuhan Yang Maha Esa karena atas rahmat-Nya sehingga kami dapat menyelesaikan Laporan Praktik Kerja Lapangan Politeknik Statistika STIS Tahun Akademik 2022/2023 dengan judul “Kajian Pemanfaatan *Mobile Positioning Data* (MPD) dalam Menunjang Statistik Pariwisata” dengan tepat waktu.

Kami mengucapkan terima kasih kepada:

1. Kepala BPS RI dan Dewan Penyantun sebagai pihak yang memberikan kami kesempatan untuk melaksanakan Praktik Kerja Lapangan (PKL),
2. Kepala BPS Provinsi Jawa Timur, Kepala BPS Kota Surabaya, Kepala BPS Kota Malang, Kepala BPS Kota Batu, Kepala BPS Kabupaten Malang, dan seluruh staf yang membantu dalam perizinan dan kelancaran pelaksanaan PKL,
3. Seluruh Dosen Pembimbing Praktik Kerja Lapangan Politeknik Statistika STIS Tahun Akademik 2022/2023,
4. Seluruh Mahasiswa Tingkat III Tahun Akademik 2021/2022 Politeknik Statistika STIS atas kerja keras, waktu, dan keikhlasannya untuk berkontribusi secara penuh dalam kegiatan PKL.

Kami berharap laporan ini dapat bermanfaat dan dapat dijadikan referensi demi pengembangan ilmu pengetahuan statistik maupun pariwisata ke arah yang lebih baik. Kami menyadari bahwa masih terdapat kekurangan dalam laporan yang kami susun. Oleh karena itu, kami mengharapkan saran dan kritik membangun sehingga dapat menjadi koreksi untuk pelaksanaan PKL mendatang.

Jakarta, 09 Juli 2023

Koordinator PKL Politeknik Statistika STIS

Tahun Akademik 2022/2023

Dr. Siti Muchlisoh, M.Si

NIP 197002191992112001



ABSTRAK

Sejak tahun 2016, Badan Pusat Statistik (BPS) telah mulai melakukan eksplorasi pemanfaatan MPD (*Mobile Positioning Data*) sebagai sumber data lain dalam penyusunan statistik pariwisata secara umum selain menggunakan Survei Wisatawan Nusantara secara konvensional. Namun, MPD sebagai salah satu sumber data juga masih memiliki kekurangan, diantaranya adalah data yang masih perlu dipastikan kualitasnya, tidak bisa merekam perjalanan wisatawan yang tidak memiliki ponsel, dan kemungkinan terjadinya *double counting* akibat penggunaan banyak kartu sim pada satu *handphone*. Oleh karena itu, sebagai solusi agar pemanfaatan data MPD lebih optimal, penelitian ini melakukan pembangunan algoritma *family grouping* dan *double counting* menggunakan data MPD yang sudah terlebih dahulu dipastikan kualitasnya melalui tahapan *quality assurance*. Algoritma H3Geo dengan *staypoint* dan resolusi 10 merupakan algoritma terbaik untuk mengidentifikasi *family grouping* menggunakan data MPD dengan F1-score sebesar 88% dan akurasi 98%. Sementara itu, algoritma terbaik untuk mengidentifikasi *double counting* pada data MPD adalah algoritma kombinasi jarak terdekat antar *staypoint* dengan skema data tahunan dengan F1-score sebesar 64% dan akurasi 98%. Untuk melengkapi karakteristik wisatawan nusantara yang tidak bisa didapatkan melalui data MPD, penelitian ini juga melakukan survei konvensional di Kota Malang dengan *time reference* triwulan IV tahun 2022. Dari hasil survei, ditemukan bahwa wisatawan nusantara asal Kota Malang didominasi oleh kelompok umur 25–49 tahun dengan mayoritas perjalanan wisata dilakukan untuk berkunjung ke objek wisata. Selain itu, ditemukan perbedaan yang signifikan antara jumlah perjalanan wisata yang tercatat pada data MPD dan pada data hasil survei konvensional.

Kata Kunci : Statistik Wisatawan Nusantara, *Mobile Positioning Data*, *Family Grouping*, *Double Counting*



DAFTAR ISI

PRAKATA.....	v
ABSTRAK.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xvi
DAFTAR LAMPIRAN.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi dan Batasan Penelitian.....	6
1.3 Rumusan Masalah.....	10
1.4 Tujuan Penelitian.....	10
1.5 Manfaat Penelitian.....	11
BAB II TINJAUAN PUSTAKA.....	12
2.1 Landasan Teori.....	12
2.1.1 Wisata & Pariwisata.....	12
2.1.2 Wisatawan.....	12
2.1.3 Wisatawan Nusantara.....	13
2.1.4 Perjalanan Wisata.....	14
2.1.5 Objek Wisata.....	15
2.1.6 Mobile Positioning Data (MPD).....	16
2.1.7 Usual Environment.....	18
2.1.8 AMDA.....	20
2.1.9 Family Grouping.....	21
2.1.10 Double Counting.....	22
2.1.11 Machine Learning.....	22
2.1.12 Python.....	23
2.1.13 Haversine Distance.....	24
2.1.14 Hexagonal Hierarchical Geospatial Indexing System (H3).....	24
2.1.15 Dynamic Time Warping.....	27
2.1.16 Frechet Distance.....	30
2.1.17 Decision Tree dan Gini Impurity.....	31



2.2 Penelitian Terkait.....	33
2.3 Kerangka Pikir.....	36
BAB III METODOLOGI.....	39
3.1 Ruang Lingkup Penelitian.....	39
3.1.1 Waktu dan Tempat Penelitian.....	39
3.1.2 Cakupan Penelitian.....	39
3.1.3 Alat Operasional.....	43
3.2 Metode Penelitian.....	47
3.2.1 Penjaminan Kualitas dan Pembangunan Algoritma Family Grouping dan Double Counting.....	47
3.2.2 Penyajian Statistik Wisatawan Nusantara melalui Pendekatan MPD dan Survei Konvensional.....	47
3.3 Metode Pengumpulan Data.....	48
3.4 Metode Penarikan Sampel.....	49
3.4.1 Desain Sampling.....	49
3.4.2 Teknik Estimasi Parameter Desain Sampling.....	51
3.4.3 Penentuan Jumlah Sampel.....	52
3.5 Metode Analisis.....	53
3.5.1 Analisis Deskriptif.....	53
3.5.2 Evaluasi Performansi.....	56
3.6 Definisi Operasional.....	68
3.6.1 Survei Kajian MPD (SPRT.PKL62).....	68
3.6.2 Survei Wisatawan Nusantara (VRTW.62.C).....	73
BAB IV HASIL DAN PEMBAHASAN.....	84
4.1 Gambaran Umum Karakteristik Relawan.....	85
4.1.1 Jumlah Relawan berdasarkan Kepemilikan Handphone Pribadi.....	85
4.1.2 Jumlah Anggota Keluarga yang Tidak Memiliki Handphone Pribadi.....	86
4.1.3 Jumlah Kartu sim Aktif yang Digunakan Relawan.....	88
4.1.4 Jumlah Relawan dengan Kepemilikan Ganda Kartu sim.....	89

4.1.5 Jumlah Relawan Family Grouping dan Double Counting.....	91
4.2 Penjaminan Kualitas Data (Data Quality Assurance).....	92
4.2.1 Distribusi Jumlah Records.....	99
4.2.2 Jumlah Records Duplicate.....	97
4.2.3 Latitude, Longitude, dan LAC Not NULL.....	99
4.2.4 Jumlah Records dan Subscribers per Hari Berdasarkan Tipe Data Source.....	101
4.2.5 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari.....	104
4.2.6 Jumlah Subscribers yang Hadir per Bulan.....	106
4.2.7 Average Record per Hour Selama Satu Tahun.....	109
4.3 Karakteristik Data Training dan Data Testing.....	111
4.4 Pembangunan Algoritma Family Grouping dan Double Counting.....	121
4.4.1 Penentuan Usual Environment Menggunakan AMDA.....	121
4.4.2 Pengelompokan Subscribers dengan Lokasi Home yang Sama.....	130
4.4.3 Eksplorasi Temuan.....	133
4.4.4 Periode Waktu Analisis.....	142
4.4.5 Eliminasi Pergerakan Tidak Wajar.....	145
4.4.6 Pembentukan Staypoint.....	148
4.4.7 Pembangunan Algoritma dengan Kombinasi Jarak Terdekat antar Staypoint.....	152
4.4.8 Pembangunan Algoritma Hexagonal Hierarchical Geospatial Indexing System (H3).....	159
4.4.9 Pembangunan Algoritma dengan Dynamic Time Warping dan Frechet Distance.....	173
4.4.10 Pembangunan Threshold.....	178
4.5 Hasil Implementasi Algoritma.....	183
4.5.1 Family Grouping.....	183
4.5.2 Double Counting.....	199
4.6 Statistik Wisatawan Nusantara Asal Kota Malang.....	216
4.6.1 Jumlah Wisatawan Nusantara Asal Kota Malang berdasarkan Kelompok Umur.....	217
4.6.2 Jumlah Wisatawan Nusantara Asal Kota Malang berdasarkan Status Kepemilikan Handphone.....	218



4.6.3 Jumlah Kartu Sim Aktif Wisatawan Nusantara berdasarkan Provider.....	221
4.6.4 Jumlah Wisatawan Nusantara dengan Kepemilikan Ganda Kartu sim Aktif dalam Handphone Utama.....	223
4.6.5 Jumlah Perjalanan Wisata Asal Kota Malang berdasarkan Kelompok Umur.....	225
4.6.6 Jumlah Perjalanan Wisata Asal Kota Malang berdasarkan Bulan.....	227
4.6.7 Persentase Perjalanan Wisata Asal Kota Malang berdasarkan Kategori Perjalanan Wisata.....	229
BAB V PENUTUP.....	231
5.1 Kesimpulan.....	231
5.2 Saran.....	233
DAFTAR PUSTAKA.....	235
LAMPIRAN.....	243

DAFTAR GAMBAR

Gambar 2.1 Sistem Pengindeksan Geospasial dengan H3	25
Gambar 2.2 Ilustrasi DTW.....	28
Gambar 2.3 Optimal Warping Path dari 2 Deret Waktu X dan Y	29
Gambar 2.4. Ilustrasi Frechet Distance.....	30
Gambar 2.5 Decision Tree.....	32
Gambar 2.6 Kerangka Pikir.....	38
Gambar 3.1 Skema Sampling.....	50
Gambar 3.2 Flowchart Algoritma Penentuan Trip.....	59
Gambar 3.3 Kategori Objek Wisata dalam Survei Kajian MPD.....	62
Gambar 3.4 Kategori Objek Wisata dalam Survei Wisatawan Nusantara.....	76
Gambar 4.1 Jumlah Relawan Berdasarkan Kepemilikan Handphone Pribadi Menurut Kelompok Umur.....	85
Gambar 4.2 Jumlah Anggota Keluarga Relawan yang Tidak Memiliki Handphone Pribadi Menurut Umur.....	87
Gambar 4.3 Jumlah Kartu sim Aktif yang Digunakan Relawan pada Handphone Utama Berdasarkan Provider.....	88
Gambar 4.4 Jumlah Relawan dengan Kepemilikan Ganda Kartu sim Aktif pada Handphone Utama Berdasarkan Kombinasi Provider.....	90
Gambar 4.5 Jumlah Relawan Family Grouping dan Double Counting.....	91
Gambar 4.6 Jumlah Nomor Family Grouping dan Double Counting.....	92
Gambar 4.7 Distribusi Jumlah Records Provider Telkomsel.....	96
Gambar 4.8 Distribusi Jumlah Records Provider Indosat.....	97
Gambar 4.9 Jumlah Records Duplicate Provider Telkomsel.....	97
Gambar 4.10 Jumlah Records Duplicate Provider Indosat.....	98
Gambar 4.11 Distribusi Jumlah Record Indosat Setelah Record Duplicate Dihapus.....	99
Gambar 4.12 Normalized Latitude, Longitude, dan LAC Not Null Provider Telkomsel.....	100
Gambar 4.13 Normalized Latitude, Longitude, dan LAC Not Null Provider Indosat.....	101
Gambar 4.14 Jumlah Records per Hari Berdasarkan Tipe Data Source Provider Telkomsel.....	102

Gambar 4.15 Jumlah Subscribers per Hari Berdasarkan Tipe Data Source Provider Telkomsel.....	102
Gambar 4.16 Jumlah Records per Hari Berdasarkan Tipe Data Source Provider Indosat.....	103
Gambar 4.17 Jumlah Subscribers per Hari Berdasarkan Tipe Data Source Provider Indosat.....	104
Gambar 4.18 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari Provider Telkomsel.....	104
Gambar 4.19 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari Provider Indosat.....	105
Gambar 4.20 Jumlah Subscribers yang Hadir Provider Telkomsel.....	107
Gambar 4.21 Jumlah Subscribers yang Hadir per Bulan Provider Indosat.....	108
Gambar 4.22 Average record per hour Provider Telkomsel.....	109
Gambar 4.23 Rata-Rata Jumlah Records per Jam Provider Indosat.....	110
Gambar 4.24 Distribusi Jumlah Records Data Training Setelah Records Duplicate dihapus.....	112
Gambar 4.25 Distribusi Jumlah Records Data Testing Setelah Records Duplicate dihapus.....	112
Gambar 4.26 Latitude, Longitude, dan LAC Not Null Data Training setelah Records Duplicate Dihapus.....	113
Gambar 4.27 Latitude, Longitude, dan LAC Not Null Data Testing Setelah Records Duplicate Dihapus.....	114
Gambar 4.28 Jumlah Records per Hari Data Training Setelah Records Duplicate Dihapus.....	115
Gambar 4.29 Jumlah Records per Hari Data Testing Setelah Records Duplicate Dihapus.....	115
Gambar 4.30 Jumlah Subscribers per Hari Data Training Setelah Records Duplicate Dihapus.....	116
Gambar 4.31 Jumlah Subscribers per Hari Data Testing Setelah Records Duplicate Dihapus.....	117
Gambar 4.32 Jumlah Subscribers Data Training yang Hadir per Bulan.....	118
Gambar 4.33 Jumlah Subscribers Data Testing yang Hadir per Bulan.....	120
Gambar 4.34 Flowchart Tahapan Penentuan Usual Environment dengan AMDA.....	122

Gambar 4.35 Sebaran Raw Data msisdn00260 (Hijau) dan msisdn00325 (Merah).....	133
Gambar 4.36 Hasil Penentuan Kandidat Lokasi Home msisdn00260 (Hijau) dan msisdn00325 (Merah) Menggunakan AMDA.....	134
Gambar 4.37 Sebaran Raw Data msisdn00288 (Hijau) dan msisdn00331 (Merah).....	135
Gambar 4.38 Hasil Penentuan Kandidat Lokasi Home msisdn00288 (Hijau) dan msisdn00331 (Merah) Menggunakan AMDA.....	135
Gambar 4.39 Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah) Rentang Waktu 1 Desember 2021 - 22 Oktober 2022.....	137
Gambar 4.40 Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah) Rentang Waktu 1 Desember 2021 - 30 November 2022.....	137
Gambar 4.41 Hasil Penentuan Kandidat Lokasi Home msisdn00019 (Hijau) dan msisdn00096 (Merah) Menggunakan AMDA.....	138
Gambar 4.42 Sebaran Raw Data msisdn00079 (Hijau) dan msisdn00080 (Merah) Rentang Waktu 01 Desember 2021 - 27 Juli 2022.....	139
Gambar 4.43 Sebaran Raw Data msisdn00079 (Hijau) dan msisdn00080 (Merah) Rentang Waktu 1 Desember 2021 - 30 November 2022.....	140
Gambar 4.44 Hasil Penentuan Kandidat Lokasi Home msisdn00079 (Hijau) dan msisdn00080 (Merah) Menggunakan AMDA.....	141
Gambar 4.45 Jarak Antar staypoint pada Lima Sampel Family Nomor Family Grouping.....	144
Gambar 4.46 Jarak Antar Staypoint pada Lima Sampel Pasangan Nomor Double Counting.....	145
Gambar 4.47 Flowchart Perhitungan Jarak antar Staypoint dengan Haversine...	152
Gambar 4.48 Flowchart Penghitungan Jumlah Irisan H3 Tanpa Staypoint.....	162
Gambar 4.49 Sebaran Heksagon msisdn00xxx1.....	162
Gambar 4.50 Sebaran Heksagon msisdn00xxx2.....	163
Gambar 4.51 Sebaran Heksagon msisdn00xxx1 dan msisdn00xxx2.....	163
Gambar 4.52 Jumlah Heksagon yang Beririsan.....	164
Gambar 4.53 Flowchart Tahapan Penghitungan Jumlah Irisan H3 dengan Staypoint.....	166
Gambar 4.54 Flowchart Tahapan Metode DTW dan Frecet Distance.....	175
Gambar 4.55 Sebaran Raw Data dari Family yang Menggunakan Kartu SIM	



Telkomsel Seluruhnya.....	184
Gambar 4.56 Sebaran Raw Data dari Family yang Menggunakan Kartu SIM Indosat Seluruhnya.....	184
Gambar 4.57 Sebaran Raw Data dari Family yang Menggunakan Kartu SIM Telkomsel dan Indosat.....	185
Gambar 4.58 Threshold dan Perbandingan Rata-Rata Jarak Kestabilan Antar Kelas pada Data Training Algoritma Family Grouping.....	187
Gambar 4.59 Hasil Implementasi Metode Kombinasi Jarak Terdekat antar Staypoint pada Data Testing Algoritma Family Grouping.....	188
Gambar 4.60 Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping.....	189
Gambar 4.61 Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing Algoritma Family Grouping.....	190
Gambar 4.62 Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping.....	191
Gambar 4.63 Hasil Implementasi Metode H3 dengan Staypoint pada Data Testing Algoritma Family Grouping.....	193
Gambar 4.64 Threshold dan Perbandingan Rata-rata Jarak DTW Antar Kelas pada Data Training Algoritma Family Grouping.....	194
Gambar 4.65 Hasil Implementasi Metode DTW pada Data Testing Algoritma Family Grouping.....	195
Gambar 4.66 Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Family Grouping.....	196
Gambar 4.67 Hasil Implementasi Metode Frechet Distance pada Data Testing Algoritma Family Grouping.....	197
Gambar 4.68 Perbandingan Algoritma Family Grouping Terbaik dari Masing-masing Metode.....	198
Gambar 4.69 Sebaran Raw Data Subscriber Double Counting dengan Kombinasi Kartu SIM Telkomsel (Merah) - Telkomsel (Hijau).....	200
Gambar 4.70 Sebaran Raw Data Subscriber Double Counting dengan Kombinasi Kartu SIM Indosat (Merah) - Indosat (Hijau).....	200
Gambar 4.71 Sebaran Raw Data Subscriber Double Counting dengan Kombinasi Kartu SIM Telkomsel (Merah) - Indosat (Hijau).....	201
Gambar 4.72 Threshold dan Perbandingan Rata-rata Jarak Kestabilan antar Kelas pada Data Training Algoritma Double Counting.....	203

Gambar 4.73 Hasil Implementasi Metode Kombinasi Jarak Terdekat antar Staypoint pada Data Testing Algoritma Double Counting.....	204
Gambar 4.74 Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting.....	205
Gambar 4.75 Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing Algoritma Double Counting.....	207
Gambar 4.76 Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting.....	208
Gambar 4.77 Hasil Implementasi Metode H3 dengan Staypoint pada Data Testing Algoritma Double Counting.....	210
Gambar 4.78 Threshold dan Perbandingan Rata-rata Jarak DTW Antar Kelas pada Data Training Algoritma Double Counting.....	211
Gambar 4.79 Hasil Implementasi Metode DTW pada Data Testing Algoritma Double Counting.....	212
Gambar 4.80 Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Double Counting.....	213
Gambar 4.81 Hasil Implementasi Metode Frechet Distance pada Data Testing Algoritma Double Counting.....	214
Gambar 4.82 Perbandingan Algoritma Double Counting Terbaik dari Masing-masing Metode.....	215
Gambar 4.83 Jumlah Wisatawan Nusantara Asal Kota Malang Berdasarkan Kelompok Umur.....	217
Gambar 4.84 Jumlah Wisatawan Nusantara Asal Kota Malang Berdasarkan Status Kepemilikan Handphone dan Kelompok Umur.....	219
Gambar 4.85 Jumlah Kepemilikan Kartu sim Aktif yang Digunakan Wisatawan Nusantara Asal Kota Malang Berdasarkan Provider.....	222
Gambar 4.86 Jumlah Wisatawan Nusantara Asal Kota Malang dengan Kepemilikan Ganda Kartu sim Berdasarkan Kombinasi Provider.....	224
Gambar 4.87 Jumlah Perjalanan Wisata Asal Kota Malang Berdasarkan Kelompok Umur.....	226
Gambar 4.88 Jumlah Perjalanan Wisata Bulanan Asal Kota Malang.....	228
Gambar 4.89 Perjalanan Wisata Asal Kota Malang Berdasarkan Kategori Perjalanan Wisata.....	230



DAFTAR TABEL

Tabel 2.1 Resolusi H3.....	25
Tabel 3.1 Skema Sampling.....	51
Tabel 3.2 Confusion Matrix.....	57
Tabel 4.1 Contoh Data MPD Telkomsel.....	92
Tabel 4.2 Contoh Data MPD Indosat.....	93
Tabel 4.3 Hasil Penjaminan Kualitas Data MPD Provider Telkomsel dan Indosat.....	110
Tabel 4.4 Contoh Record MPD Sebelum Dilakukan Pembulatan.....	122
Tabel 4.5 Contoh Record MPD Setelah Dilakukan Pembulatan.....	123
Tabel 4.6 Contoh Record MPD dengan Penghitungan Haversine Distance.....	124
Tabel 4.7 Contoh Record MPD dengan Penghitungan Kecepatan.....	125
Tabel 4.8 Contoh Agregasi Penghitungan Anchor 1.....	128
Tabel 4.9 Contoh Agregasi Penghitungan Anchor 2.....	129
Tabel 4.10 Prediksi Lokasi Usual Environment.....	130
Tabel 4.11 Contoh Pengelompokan Subscribers dengan Lokasi Home yang Sama.....	131
Tabel 4.12 Prediksi Lokasi Home msisdn00260 dan msisdn00325 dengan AMDA.....	134
Tabel 4.13 Prediksi Lokasi Home msisdn00288 dan msisdn00331 dengan AMDA.....	136
Tabel 4.14. Prediksi Lokasi Home msisdn00019 dan msisdn00096 dengan AMDA.....	139
Tabel 4.15 Prediksi Lokasi Home msisdn00079 dan msisdn00080 dengan AMDA.....	141
Tabel 4.16 Data MPD dengan Data Pergerakan yang Tidak Wajar.....	147
Tabel 4.17 Contoh Pembentukan Staypoint.....	149
Tabel 4.18 MSISDN 00xxx1 dengan 2 Staypoint pada Interval 1.....	153
Tabel 4.19 MSISDN 00xxx2 dengan 3 Staypoint pada Interval 1.....	155
Tabel 4.20 Contoh Rata-Rata Kestabilan Jarak Terdekat antar Nomor (MSISDN).....	158
Tabel 4.21 Contoh Konversi Koordinat Lokasi Menjadi Indeks H3 (Resolusi 8).....	160
Tabel 4.22 Contoh Indeks H3 yang Unik.....	161
Tabel 4.23 Jumlah Heksagon yang Beririsan antar Nomor (MSISDN).....	164

Tabel 4.24 Contoh Penghitungan Irisan Indeks H3 Tanpa Pembentukan Staypoint msisdn00xxx1.....	167
Tabel 4.25 Contoh Penghitungan Irisan Indeks H3 Tanpa Pembentukan Staypoint msisdn00xxx2.....	168
Tabel 4.26 Contoh Penghitungan Irisan Indeks H3 dengan Pembentukan Staypoint msisdn00xxx1.....	169
Tabel 4.27 Contoh Penghitungan Irisan Indeks H3 dengan Pembentukan Staypoint msisdn00xxx2.....	170
Tabel 4.28 Rata-rata Jumlah Heksagon yang Beririsasi per Hari antar Nomor (MSISDN).....	172
Tabel 4.29 Lokasi Home msisdn00xxx Untuk Setiap Bulan.....	176
Tabel 4.30 Contoh Jarak DTW antar Nomor (MSISDN).....	177
Tabel 4.31 Contoh Jarak Frechet antar Nomor (MSISDN).....	177
Tabel 4.32 Contoh Dataset Rata-rata Kestabilan Jarak Terdekat antar Nomor (MSISDN).....	178
Tabel 4.33 Pengurutan Berdasarkan Nilai Average_D dan Hasil Penghitungan rata-rata Average_D.....	180
Tabel 4.34 Gini Impurity untuk Setiap Threshold Rata-rata Average_D.....	181
Tabel 4.35 Classification Report Algoritma H3 dengan Staypoint Resolusi 10.....	199
Tabel 4.36 Classification Report Algoritma Kombinasi Jarak Terdekat antar Staypoint (Tahunan).....	216



DAFTAR LAMPIRAN

Lampiran 1. Tabulasi Hasil Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint untuk Family Grouping.....	243
Lampiran 2. Tabulasi Hasil Running Algoritma H3 Tanpa Pembentukan Staypoint untuk Family Grouping.....	243
Lampiran 3. Tabulasi Hasil Running Algoritma H3 dengan Pembentukan Staypoint untuk Family Grouping.....	246
Lampiran 4. Tabulasi Hasil Running Dynamic Time Warping untuk Family Grouping.....	249
Lampiran 5. Tabulasi Hasil Running Frechet Distance untuk Family Grouping.....	250
Lampiran 6. Tabulasi Hasil Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint untuk Double Counting Periode Bulanan.....	250
Lampiran 7. Tabulasi Hasil Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint untuk Double Counting Periode Tahunan.....	250
Lampiran 8. Tabulasi Hasil Running Algoritma H3 Tanpa Pembentukan Staypoint untuk Double Counting.....	251
Lampiran 9. Tabulasi Hasil Running Algoritma H3 dengan Pembentukan Staypoint untuk Double Counting.....	254
Lampiran 10. Tabulasi Hasil Running Dynamic Time Warping untuk Double Counting.....	257
Lampiran 11. Tabulasi Hasil Running Frechet Distance untuk Double Counting.....	257
Lampiran 12. Kuesioner Survei Kajian MPD (SPRT.PKL62).....	258
Lampiran 13. Kuesioner Survei Wisatawan Nusantara (VRTW.62.C).....	263
Lampiran 14. Agreement Relawan Kajian MPD Provider Telkomsel.....	267
Lampiran 15. Agreement Relawan Kajian MPD Provider Indosat.....	268

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia adalah negara yang kaya akan keanekaragaman hayati dan sumber daya alam. Indonesia memiliki sekitar 16.766 pulau (Kementerian Dalam Negeri, 2021) dan ribuan suku dengan latar budaya yang beragam. Selain itu, Indonesia juga memiliki pemandangan alam yang sangat indah sehingga membuat Indonesia menjadi negara yang memiliki daya tarik tersendiri untuk dikunjungi, baik oleh wisatawan domestik maupun wisatawan mancanegara. Hal tersebut membuat Indonesia memiliki potensi besar dalam mengembangkan sektor pariwisatanya. Hal tersebut terbukti dengan jumlah usaha Objek Daya Tarik Wisata (ODTW) Indonesia pada tahun 2021 tercatat sebanyak 2.563 usaha (Statistik Objek Daya Tarik Wisata 2021, 2022) dengan jumlah perjalanan wisatawan nusantara mencapai 62,8 juta pada bulan Oktober 2021 (Statistik Wisatawan Nusantara 2021, 2022).

Sektor pariwisata di Indonesia menjadi sektor yang dapat memberikan kontribusi pada perekonomian nasional. Hal ini dibuktikan dengan besarnya kontribusi pariwisata terhadap Produk Domestik Bruto (PDB). Berdasarkan data dari Badan Pusat Statistik (BPS), proporsi kontribusi pariwisata terhadap PDB tahun 2017 adalah sebesar 4,11 persen. Sektor pariwisata diharapkan menjadi penggerak utama dalam menciptakan lapangan pekerjaan, usaha, penerimaan devisa, serta pembangunan infrastruktur guna mempercepat pertumbuhan ekonomi Indonesia.

Sebelum pandemi Covid-19, pemerintah gencar dalam menarik wisatawan mancanegara untuk berkunjung ke Indonesia dengan cara memperkuat promosi wisata, mendorong investasi, dan mempercepat penyelesaian proyek infrastruktur dalam rangka meningkatkan devisa. Selain gencar dalam menarik wisatawan mancanegara, pemerintah juga terus meningkatkan jumlah kunjungan wisatawan



nusantara (wisnus). Hal ini dapat terlihat dari data BPS yang menunjukkan bahwa terjadi peningkatan jumlah perjalanan wisnus di Indonesia yang semula sebanyak 303.403.888 perjalanan di tahun 2018 menjadi 722.158.733 perjalanan di tahun 2019. Namun sejak covid menyerang pada tahun 2020, sektor pariwisata menjadi terpukul dan bahkan mati suri. Data BPS menunjukkan bahwa jumlah perjalanan wisata pada tahun 2020 hanya sebesar 524.571.392, mengalami penurunan dari tahun sebelumnya. Oleh karena itu, perlu ada usaha untuk mengembalikan perkembangan sektor pariwisata.

Dalam memonitoring perkembangan sektor pariwisata Indonesia, dibutuhkan ketersediaan data, informasi, serta semua aspek terkait kunjungan wisatawan secara rinci, cermat, dan tepat waktu. Untuk itu, diperlukan rangkaian kegiatan pengumpulan data wisatawan yang terencana dan berkesinambungan (Badan Pusat Statistik, 2021). Sejak tahun 1981 sampai saat ini, Badan Pusat Statistik telah melakukan kegiatan pengumpulan data wisatawan nusantara dan hasilnya dipublikasikan melalui publikasi Statistik Wisatawan Nusantara (Badan Pusat Statistik, 2023). Publikasi ini merupakan publikasi rutin tahunan yang disusun dan disajikan oleh BPS yang berisi tentang profil wisatawan nusantara, maksud/tujuan perjalanan, akomodasi yang digunakan, lama perjalanan, dan rata-rata pengeluaran terkait dengan perjalanan yang dilakukan oleh penduduk Indonesia. Sejak saat itu juga, statistik wisnus yang ada di publikasi rutin tersebut didapatkan melalui kegiatan Survei Wisatawan Nusantara secara konvensional. Namun sayangnya, kegiatan survei ini memiliki sejumlah keterbatasan. Survei hanya dapat digunakan untuk mengestimasi tingkat provinsi dan hasilnya bergantung pada daya ingat responden. Hal ini berpotensi menimbulkan kesalahan dari sisi jawaban responden dan kesalahan dalam pengambilan sampel survei (BPS, 2021).

Untuk mengatasi kelemahan tersebut, sejak tahun 2016, Badan Pusat Statistik (BPS) telah mulai melakukan eksplorasi pemanfaatan MPD (*Mobile Positioning Data*) sebagai sumber data lain dalam penyusunan statistik pariwisata secara umum. MPD merupakan suatu dataset berskala besar mengenai catatan

transaksi dan lokasi dari operator seluler (*Mobile Network Operator/MNO*) yang diproses dan disimpan di dalam suatu sistem (Ahas et al., 2010). Penggunaan MPD dipilih karena memiliki beberapa keunggulan (Ahas et al., 2010), yaitu:

1. Penggunaan telepon seluler tersebar luas dan populer, baik di negara maju dan berkembang. Saat ini, hampir seluruh masyarakat memiliki telepon seluler dan tidak terbatas pada kalangan tertentu saja;
2. Kecenderungan masyarakat untuk selalu membawa telepon seluler dan menjadikannya barang penting;
3. Data awal yang berbentuk ialah digital sehingga bebas dari *human error* seperti keterbatasan ingatan responden atau kesalahan entri data;
4. Penggunaan MPD memungkinkan dalam mempelajari perpindahan penduduk dalam dimensi ruang dan waktu secara lebih tepat.

Sebagai salah satu sumber *big data*, MPD memiliki masalah terkait karakteristik *veracity*, yaitu perihal kualitas, keakuratan, dan ketidakpastian data (Munaf et al., 2021). Sebelum melakukan pemrosesan lebih lanjut, penting untuk melakukan kontrol kualitas awal dan pembersihan data MPD. Tahapan tersebut diantaranya adalah menghapus *duplicate*, memeriksa kualitas data dan koordinat BTS, mendeteksi data yang hilang, dsb. (Saluveer et al., 2020). Menurut Saluveer et al. (2020), MNO menggunakan satuan milidetik sebagai unit pengukuran dan terkadang mengumpulkan informasi yang sama beberapa kali dalam *database*. Oleh karena itu, untuk pemrosesan yang lebih baik, maka catatan duplikat ini perlu dieliminasi. Untuk tujuan penjaminan kualitas, data MPD perlu diperiksa sesuai dengan format yang disepakati untuk mengidentifikasi kesalahan dan mengevaluasi apakah data yang ada merepresentasikan nilai-nilai yang logis. Sebagai contoh, koordinat BTS harus berada dalam batas negara/area tertentu yang diharapkan (Saluveer et al., 2020). Kemudian, terkait data MPD yang hilang, Saluveer et al. (2020) menjelaskan terdapat empat jenis data hilang yang teridentifikasi, salah satunya adalah data yang hilang pada area tertentu (satu atau lebih BTS) dalam periode tertentu. Data yang hilang ini harus dikonfirmasi dan diminta kembali dari MNO (Saluveer et al., 2020). Berdasarkan penjelasan



tersebut, penting untuk memeriksa apakah data MPD memiliki kualitas yang cukup baik dan dapat digunakan untuk pemrosesan analisis lebih lanjut dengan melakukan penjaminan kualitas (*Quality Assurance*).

Tingginya penggunaan telepon seluler dalam jumlah, luas, dan data yang dihasilkan secara terus menerus dapat dianalisis lebih lanjut dalam penelitian yang melibatkan pergerakan manusia (Ahas et al., 2007). Di Indonesia, jumlah pelanggan telepon seluler berdasarkan data Statistik Telekomunikasi Indonesia yang dirilis oleh BPS pada tahun 2021 adalah sebesar 65,87%. Proporsi ini mengalami peningkatan sebesar 3,03% dari tahun sebelumnya. Hal ini menunjukkan adanya peningkatan dari proporsi masyarakat yang membawa telepon genggam kapanpun dan dimanapun. Situasi ini menjadi salah satu alasan pentingnya mempelajari pergerakan manusia melalui data yang dihasilkan oleh telepon seluler.

Menurut data BPS pada tahun 2022, proporsi individu yang memiliki telepon genggam pada kelompok umur kurang dari 15 tahun adalah sekitar 40,25%, pada kelompok umur 15-24 tahun sekitar 91,82 %, pada kelompok umur 25-64 tahun sekitar 74,09%, dan pada kelompok umur lebih dari 65 tahun sekitar 27,46%. Statistik ini menunjukkan bahwa tidak semua penduduk memiliki telepon genggam. Jika perjalanan wisata dihitung menggunakan MPD, perjalanan individu yang tidak memiliki telepon genggam tidak dapat direkam. Pendekatan lain yang dapat dilakukan adalah dengan mengestimasi perjalanan wisata yang dilakukan bersama orang terdekatnya, seperti keluarga. Hal tersebut sejalan dengan ditemukannya fenomena bahwa banyak orang cenderung melakukan perjalanan wisata bersama keluarga atau kerabat terdekat. Hal tersebut didukung juga oleh penelitian Koranti, Sriyanto, dan Sidik (2017) tentang Analisis Preferensi Wisatawan Terhadap Sarana di Wisata Taman Wisata Kopeng. Penelitian tersebut menyimpulkan bahwa kedatangan wisatawan bersama teman/kerabat merupakan kelompok yang paling banyak dijumpai di Taman Wisata Kopeng diikuti dengan wisatawan yang datang bersama keluarga, yaitu sebesar 39.68%. Kelompok keluarga ini merupakan wisatawan yang datang

bersama anak, istri, suami, atau keluarga besar mereka. Oleh karena itu, penentuan individu ke dalam suatu keluarga (*family grouping*) berdasarkan *usual environment* dan pola perjalanan menggunakan MPD menarik untuk dilakukan.

Selain terkait *veracity*, terdapat tantangan lain dalam menganalisis dan mendapatkan jumlah perjalanan wisata dengan MPD, yaitu data *mobile positioning* seseorang yang dihasilkan oleh operator dapat berbeda meskipun orang tersebut berada di lokasi dan waktu yang sama. Hal ini dipengaruhi oleh jangkauan dan lokasi tower seluler (BTS) tiap operator serta aktivitas seluler untuk nomor masing-masing. Jika seorang wisatawan menggunakan dua nomor berbeda dalam satu ponsel sebagai penentu *mobile positioning*, wisatawan tersebut akan terhitung sebanyak dua kali. Selain itu, jumlah perjalanan wisata yang dilakukan wisatawan tersebut juga akan terhitung sebanyak dua kali. Hal inilah yang kemudian disebut dengan *double counting* pada data MPD. Untuk mengatasi hal ini, diperlukan metode atau penanganan terhadap data MPD yang sebenarnya merepresentasikan satu individu namun seolah-olah terhitung layaknya dua individu (*double counting*), sehingga data MPD tersebut dapat menentukan posisi dan aktivitas wisatawan secara akurat untuk kepentingan lebih lanjut.

Seperti yang telah dijabarkan sebelumnya, pemanfaatan MPD sebagai sumber data alternatif dalam penyusunan statistik pariwisata secara umum merupakan potensi yang perlu dikembangkan. Salah satu indikator statistik pariwisata yang dapat diestimasi menggunakan MPD adalah jumlah perjalanan wisata (BPS, 2021). Dibalik kelebihan yang ditawarkannya, MPD juga memiliki kelemahan dalam mengumpulkan data wisatawan. Salah satunya adalah informasi mengenai karakteristik pengguna yang tidak dapat diperoleh jika hanya menggunakan MPD tanpa melakukan survei (Ahas et al., 2009). Selain itu, MPD juga tidak bisa menangkap jumlah perjalanan wisata yang tidak melibatkan penggunaan telepon seluler. Beberapa kekurangan MPD tersebut menjadi cukup merugikan mengingat data wisata yang komprehensif dibutuhkan dalam hal pengembangan wisata guna memulihkan sektor pariwisata. Oleh karena itu,

diperlukan pengumpulan data lain untuk melengkapi data wisatawan nusantara yang belum mampu dihasilkan oleh pendekatan MPD melalui survei konvensional.

Pada penelitian ini, akan dilakukan penjaminan kualitas (*Quality Assurance*) pada data MPD sebelum dilakukan pemrosesan lebih lanjut untuk kebutuhan analisis. Kemudian, akan dilakukan pengelompokan individu yang termasuk dalam satu keluarga (*family grouping*) berdasarkan *usual environment* dan pola perjalanan wisata melalui MPD. Selain itu, penelitian ini akan mencoba mengatasi permasalahan *double counting* pada data MPD di dua nomor berbeda milik Telkomsel dan Indosat dengan cara menganalisis pola dari kedua nomor tersebut. Terakhir, akan disajikan statistik wisatawan nusantara terutama seputar perjalanan wisata menggunakan pendekatan MPD dan survei konvensional.

1.2 Identifikasi dan Batasan Penelitian

Fokus utama dari penelitian ini adalah terkait pemanfaatan dan penanganan MPD (*mobile positioning data*) agar data yang dihasilkan cukup baik dan layak untuk diolah, sehingga output atau estimasi yang dihasilkan menjadi lebih akurat. Hal yang pertama kali dilakukan terkait penanganan dan upaya perbaikan kualitas data MPD adalah dengan melakukan penjaminan kualitas (*quality assurance*). Setelah data MPD dapat dipastikan memiliki kualitas yang cukup baik untuk digunakan, penelitian ini mencoba memanfaatkan data MPD tersebut untuk mengidentifikasi individu yang termasuk ke dalam satu keluarga (*family*) dan juga mengidentifikasi individu-individu yang termasuk sebagai entitas *double counting* yang dilihat berdasarkan *usual environment* dan pola perjalanan yang dilakukan.

Dalam MPD, *family grouping* dapat diartikan sebagai pengelompokan keluarga berdasarkan analisis *usual environment* dan pola perjalanan dari data MPD. Analisis ini penting untuk dilakukan mengingat MPD tidak mencakup wisatawan yang tidak memiliki ponsel. Selain itu, terdapat sebuah fenomena bahwa sebagian besar perjalanan seseorang dilakukan bersama keluarga atau kerabat terdekat (Koranti et al., 2017) sehingga pendekatan yang dapat dilakukan

adalah dengan mengestimasi perjalanan wisata yang dilakukan bersama orang terdekatnya, seperti keluarga.

Double counting dapat diartikan sebagai individu yang sama dalam kepemilikan dua nomor telepon yang berbeda dalam satu perangkat ponsel. Kondisi ini menyebabkan individu tersebut mempunyai *record* MPD yang berbeda antara kedua nomor sehingga seolah-olah dua *record* MPD berbeda tersebut merepresentasikan aktivitas seluler dari dua individu yang berbeda juga. Chaplove (2017) menjelaskan bahwa penghitungan ganda (*double counting*) harus dihindari karena meningkatkan jumlah orang yang dicapai. Adapun menurut Singh (2020) *double counting* akan mengakibatkan *overestimate* pada sebuah estimator, sehingga estimator tersebut tidak layak digunakan sebagai penaksir populasi. Dalam ranah pariwisata, *double counting* menyebabkan seorang wisatawan terhitung sebanyak dua kali, atau jumlah perjalanan wisata yang dilakukan terhitung sebanyak dua kali. Oleh karena itu, *double counting* sebaik mungkin dihindari dengan cara menerapkan metode yang dapat mengatasi permasalahan tersebut.

Untuk memperoleh data MPD, penelitian bekerja sama dengan Badan Pusat Statistik (BPS) yang telah berkolaborasi dengan PT Telekomunikasi Selular (Telkomsel) dan PT Indosat Tbk (Indosat Ooredoo Hutchison). Telkomsel dan Indosat merupakan dua dari lima perusahaan operator seluler yang memiliki *marketplace* terbesar di Indonesia. Hal tersebut dapat dibuktikan dalam penelitian Octasylva dan Rurianto (2020) yang menghasilkan Data *Market Share* Industri Telekomunikasi dimana PT Telekomunikasi Selular (Telkomsel) menempati urutan pertama sebagai perusahaan operator seluler terbesar di Indonesia dengan jumlah pelanggan sebanyak 168,6 juta dan menguasai sebesar 52,97% pangsa pasar. Sedangkan PT Indosat Tbk (Indosat) menempati urutan ketiga dengan jumlah pelanggan sebanyak 53,3 juta pelanggan dan menguasai sekitar 16,75% pangsa pasar.

Kedua perusahaan layanan telekomunikasi tersebut menyediakan *record* data MPD milik penggunanya. Untuk mendapatkan akses penuh terhadap data



MPD yang digunakan dalam penelitian ini, diperlukan relawan, baik pengguna Telkomsel maupun Indosat, yang bersedia untuk memberikan data MPD-nya melalui sebuah dokumen perjanjian tertulis. Demi kemudahan dan kelancaran proses pencarian relawan, perlu dilakukan pembatasan terkait siapa saja yang ditargetkan untuk menjadi relawan sekaligus unit analisis pada penelitian ini, diantaranya adalah sebagai berikut:

1. Mahasiswa Tingkat III Politeknik Statistika STIS Tahun Akademik 2022/2023 beserta anggota rumah tangganya;
2. Mahasiswa Tingkat I, II, dan IV Politeknik Statistika STIS Tahun Akademik 2022/2023 yang berdomisili di DKI Jakarta beserta anggota rumah tangganya;
3. Dosen Politeknik Statistika STIS beserta anggota rumah tangganya;
4. Pegawai BPS Kota Surabaya dan BPS Provinsi Jawa Timur beserta anggota rumah tangganya;
5. Mahasiswa Tingkat I, II, dan IV Politeknik Statistika STIS Tahun Akademik 2022/2023 yang berdomisili di Provinsi Jawa Timur beserta anggota rumah tangganya.

Dalam penelitian ini, periode yang digunakan dalam analisis data MPD adalah selama satu tahun, yaitu Januari hingga Desember 2022. Hal ini menyesuaikan dengan konsep BPS dalam mengidentifikasi *usual environment*, yaitu pengamatan *monthly home* dan *work* dalam jangka waktu 1 tahun (BPS, 2021). Oleh karena itu, dari keseluruhan daftar relawan di atas, relawan yang *eligible* dalam penelitian ini adalah relawan yang menggunakan provider Telkomsel dan atau Indosat selama periode waktu penelitian.

Selain periode penelitian, batas yang perlu ditentukan adalah terkait kode referensi yang digunakan dalam data MPD milik Telkomsel maupun Indosat. Kode referensi yang direpresentasikan oleh *Local Administrative Code* (LAC) merujuk kepada kode wilayah kerja statistik BPS tahun 2020 semester 2 atau dengan kata lain masih menggunakan kode referensi 34 provinsi (sebelum adanya pemekaran menjadi 38 provinsi pada tahun 2022). Untuk mengurangi

kompleksitas pengolahan data dan penyederhanaan validasi individu yang termasuk *family grouping* dan *double counting*, penelitian ini mengasumsikan bahwa setiap individu hanya memiliki dan menggunakan satu *handphone* yang kemudian disebut dengan handphone utama.

Survei konvensional yang dilakukan pada penelitian ini berfokus pada kegiatan wisata yang terjadi di Kota Malang, Jawa Timur. Berdasarkan Kepala Disporapar Kota Malang, Ida Ayu Made Wahyuni, jumlah kunjungan wisatawan Kota Malang yang tercatat hingga Agustus 2022 ada sebanyak lebih dari 6 juta orang dimana jumlah ini melebihi target, yaitu sebanyak 1,9 juta orang (Kemenparekraf, 2022). Selain itu, Kota Malang juga memiliki 78 daya tarik wisata yang terbagi dalam berbagai jenis wisata, diantaranya adalah wisata belanja sebanyak 22 lokasi (28%), wisata warisan arsitektur sebanyak 13 lokasi (17%), dan sisanya sebesar 55% terdiri dari berbagai macam lokasi wisata, seperti wisata monumen, museum, taman rekreasi, candi, spiritual, makam, olahraga, boulevard, dan kawasan (Mirajanatin et al., 2013). Berdasarkan hal-hal tersebut, dipilihlah Kota Malang sebagai lokus penelitian.

Periode amatan pada survei konvensional dibatasi pada rentang Oktober hingga Desember 2022 (Triwulan IV). Triwulan terakhir tahun 2022 dipilih karena keterbatasan pengumpulan data yang dilakukan secara konvensional serta adanya batasan dari kemampuan daya ingat tiap responden. Selain itu, triwulan terakhir dipilih karena dianggap sudah mengakomodasi adanya musim ramai (*peak season*) pada bulan Desember dan musim sepi (*low season*) pada bulan Oktober dan November. Oleh karena itu, unit analisis dari survei wisatawan nusantara ini adalah anggota rumah tangga di Kota Malang yang pernah melakukan perjalanan wisata selama Triwulan IV 2022.



1.3 Rumusan Masalah

Adapun rumusan masalah yang menjadi pokok bahasan penelitian ini adalah sebagai berikut:

1. Bagaimana kualitas *Mobile Positioning Data* (MPD) yang dihasilkan dari nomor-nomor relawan?
2. Bagaimana algoritma untuk mengelompokkan individu yang termasuk dalam satu keluarga berdasarkan *usual environment* dan pola perjalanan wisata (*family grouping*)?
3. Bagaimana algoritma untuk mengidentifikasi *double counting* berdasarkan pola perjalanan individu yang menggunakan lebih dari satu nomor Telkomsel dan/atau Indosat dalam satu perangkat?
4. Bagaimana hasil statistik wisatawan nusantara melalui pendekatan *Mobile Positioning Data* (MPD) dan survei konvensional di Kota Malang selama Triwulan IV 2022?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini yaitu:

1. Menerapkan penjaminan kualitas (*quality assurance*) *Mobile Positioning Data* (MPD) yang dihasilkan dari nomor-nomor relawan.
2. Membangun algoritma untuk mengelompokkan individu yang termasuk dalam satu keluarga berdasarkan *usual environment* dan pola perjalanan wisata (*family grouping*).
3. Membangun algoritma untuk mengidentifikasi *double counting* berdasarkan pola perjalanan individu yang menggunakan lebih dari satu nomor Telkomsel dan/atau Indosat berbeda dalam satu perangkat.
4. Menyajikan statistik wisatawan nusantara melalui pendekatan *Mobile Positioning Data* (MPD) dan survei konvensional di Kota Malang selama Triwulan IV 2022.



1.5 Manfaat Penelitian

Berdasarkan tujuan tersebut, penelitian ini diharapkan dapat memberikan manfaat kepada beberapa pihak, antara lain:

1. Bagi mahasiswa Politeknik Statistika STIS, sebagai salah satu sarana untuk mengimplementasikan ilmu dari mata kuliah yang sudah dipelajari di Politeknik Statistika STIS melalui Praktik Kerja Lapangan (PKL).
2. Bagi Badan Pusat Statistik (BPS), dapat membantu dalam penanganan *Mobile Positioning Data* (MPD) agar data yang dihasilkan cukup baik dan layak untuk diolah. Selain itu, juga menambah kajian dan pertimbangan mengenai penyajian statistik wisatawan nusantara melalui pendekatan MPD dan survei konvensional.
3. Bagi peneliti dan akademisi, dapat digunakan sebagai referensi untuk penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Wisata & Pariwisata

Berdasarkan Pasal 1 ayat (1) Undang-Undang Nomor 10 tahun 2009 tentang Kepariwisataan, wisata adalah kegiatan perjalanan yang dilakukan oleh seseorang atau sekelompok orang dengan mengunjungi tempat tertentu untuk tujuan rekreasi, pengembangan pribadi, atau mempelajari keunikan daya tarik wisata yang dikunjungi dalam jangka waktu sementara. Sedangkan berbagai macam kegiatan wisata yang didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah, dan pemerintah daerah disebut pariwisata. Pariwisata adalah fenomena sosial, budaya dan ekonomi yang berkaitan dengan perpindahan orang ke tempat-tempat di luar tempat tinggal mereka yang biasa dengan motivasi untuk mencari kesenangan pada umumnya (Nations, 2010). Sedangkan menurut UNWTO, pariwisata dapat diartikan sebagai aktivitas yang dilakukan oleh orang-orang yang melakukan kegiatan wisata untuk mengunjungi dan menetap di luar lingkungan asal orang tersebut untuk sementara waktu dan untuk kebutuhan kesenangan, bisnis, dan kebutuhan lain.

Berdasarkan pengertian yang telah disebutkan, dapat disimpulkan bahwa wisata adalah kegiatan perjalanan yang dilakukan oleh seseorang atau sekelompok orang dari daerah asal menuju daerah tertentu dan bersifat sementara untuk tujuan tertentu selain memperoleh penghasilan. Sedangkan pariwisata adalah penyedia atau pendukung berlangsungnya wisata, seperti pengadaan museum, usaha biro perjalanan, informasi pariwisata, konsultan, dan lain sebagainya.

2.1.2 Wisatawan

Ketika membicarakan pariwisata, tidak akan terlepas kaitannya dengan pelakunya yaitu wisatawan. Wisatawan merupakan fenomena sosial, budaya, dan ekonomi seseorang yang melakukan perjalanan kurang dari satu tahun ke suatu



tempat selain tempat tinggalnya dengan tujuan utama tertentu seperti bisnis, liburan, atau tujuan pribadi lainnya. Seseorang tidak dikatakan sebagai wisatawan ketika ia berkunjung dengan tujuan bisnis ketika ia bekerja dengan penduduk setempat dari wilayah yang dikunjungi (UNWTO, 2008). Wisatawan adalah seorang pengunjung yang menetap di suatu tempat yang dikunjungi minimal 24 jam tetapi tidak lebih dari dua belas bulan dengan tujuan pribadi (liburan, beristirahat, mengunjungi kerabat, pendidikan, pelatihan, dll) dan bisnis (konferensi, pameran, seminar, kegiatan jual beli, dll) (Badan Pusat Statistik, 2021).

Wisatawan menurut negara asalnya dibagi menjadi dua, yakni wisatawan mancanegara dan wisatawan nusantara. Wisatawan Mancanegara (wisman) adalah seorang pengunjung di suatu negara selain lingkungan tinggal biasanya, dengan keperluan tertentu selain untuk memperoleh penghasilan di negara yang dikunjungi dengan lamanya kunjungan paling lama dua belas bulan (Badan Pusat Statistik, 2021). Turis dan pelancong termasuk ke dalam wisatawan mancanegara. Sedangkan, wisatawan nusantara (wisnus) adalah suatu perjalanan yang dilakukan oleh penduduk suatu negara di dalam negaranya sendiri selain di lingkungan biasanya ia melakukan aktivitas sehari-hari dalam kurun waktu kurang dari dua belas bulan dengan tujuan selain untuk bekerja dan sekolah. Lingkungan biasa yang dimaksudkan yaitu suatu tempat yang memiliki jarak berdekatan maupun berjauhan dimana seseorang tersebut melakukan rutinitas hidup seperti biasanya secara rutin (Nations, 2010).

2.1.3 Wisatawan Nusantara

Wisatawan Nusantara (wisnus) atau *domestic visitors* adalah penduduk yang melakukan perjalanan di wilayah teritorial suatu negara, dalam hal ini Indonesia, dengan lama perjalanan kurang dari dua belas bulan dan bukan bertujuan untuk memperoleh penghasilan di tempat yang dikunjungi serta bukan merupakan perjalanan dalam rangka bekerja atau sekolah secara rutin (Badan Pusat Statistik, 2021). Ketika seorang pengunjung melakukan perjalanan di negara tempat tinggalnya, maka termasuk wisatawan nusantara atau *domestic visitors* dan



kegiatannya merupakan bagian dari pariwisata domestik (Nations & Unies, 2007). Seorang pelancong atau turis dapat dianggap sebagai wisatawan nusantara apabila memenuhi kriteria berikut:

- a. Tempat (atau wilayah) yang dikunjungi harus berada diluar *visitor's usual environment* (lingkungan yang biasa dikunjungi). Dalam hal ini tidak termasuk para pelancong yang bepergian secara teratur (setiap hari atau setiap minggu) antara tempat mereka tinggal dan tempat kerja atau bersekolah, atau sering mengunjungi tempat-tempat di dalam rutinitas kehidupan, misalnya rumah teman atau kerabat, pusat perbelanjaan, keagamaan, perawatan kesehatan atau fasilitas lain yang mungkin berada pada jarak yang cukup jauh tetapi secara teratur dan sering dikunjungi baik dalam jarak yang dekat maupun jauh;
- b. Tinggal, atau niat untuk tinggal, di tempat (atau wilayah) yang dikunjungi tidak boleh lebih dari dua belas bulan. Apabila tinggal lebih dari dua belas bulan maka tempat itu akan menjadi bagian dari *usual environment*. Kriteria ini harus diterapkan tanpa pengecualian dan juga mencakup para siswa dan pasien jangka panjang, meskipun masa inap mereka mungkin diselingi oleh tinggal sebentar di tempat asalnya atau di tempat lain;
- c. Tujuan utama kunjungan harus selain dipekerjaan oleh penduduk di tempat yang dikunjungi (Nations & Unies, 2007).

2.1.4 Perjalanan Wisata

Perjalanan Wisata adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 12 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/ gaji dari penduduk di tempat yang dituju) atau sekolah secara rutin. Tempat di luar lingkungan sehari-hari dimaksud meliputi (Badan Pusat Statistik, 2021) :

1. Berwisata ke objek wisata, dan atau
2. Mengunjungi kabupaten/kota lain dan tinggal minimal 6 jam.

Kegiatan rutin yang dimaksud adalah kegiatan sekolah dan/atau bekerja (memperoleh upah/gaji sesuai tugas pokoknya dari penduduk di tempat yang dituju), yang dilakukan secara rutin, baik frekuensi, lokasi, maupun kegiatannya. Termasuk kegiatan rutin jika mengunjungi kabupaten/kota yang sama minimal 4 kali dalam sebulan (Badan Pusat Statistik, 2021).

Sedangkan menurut IRTS, perjalanan wisata merupakan perjalanan keluar dari lingkungan sehari-hari selama kurang dari satu tahun dengan tujuan utamanya bukan untuk bekerja atau memperoleh pendapatan. Tujuan utama perjalanan dapat didefinisikan sebagai tempat yang dijadikan dasar dilakukannya perjalanan tersebut. Tujuan utama perjalanan juga dapat didefinisikan sebagai tempat dimana paling banyak waktu dihabiskan selama perjalanan atau tempat terjauh dari tempat tinggal biasanya.

2.1.5 Objek Wisata

Objek wisata adalah segala sesuatu yang ada di daerah tujuan wisata yang merupakan daya tarik agar orang-orang mau datang berkunjung ke tempat tersebut, baik yang dikelola secara komersial maupun non komersial. Contoh: gunung, danau, sungai, pantai, laut, museum, benteng, situs peninggalan sejarah, dll (Sirusa, BPS). Menurut Klasifikasi Baku Lapangan Usaha Indonesia (KBLI 2015), objek wisata dibagi menjadi beberapa macam, yaitu (Badan Pusat Statistik, 2019) :

1. Museum yang dikelola pemerintah / *Government-run museum*,
2. Museum yang dikelola swasta / *Private-run museum*,
3. Peninggalan sejarah yang dikelola pemerintah / *Historical heritage managed by the government*,
4. Peninggalan sejarah yang dikelola swasta / *Historical heritage managed privately*,
5. Taman Budaya / *Cultural Park*,
6. Pemandian Alam / *Natural Baths*,
7. Wisata Gua / *Cave Tourism*,



8. Taman Konservasi Alam / *Nature Conservation Park*,
9. Taman Nasional (TN) / *National Parks*,
10. Taman Hutan Raya (TAHURA) / *Forest Park*,
11. Taman Wisata Alam (TWA) / *Nature Park*,
12. Suaka Marga Satwa (SM) dan Cagar Alam (CA) / *Wildlife and Nature Reserves*,
13. Hutan Lindung / *Protected forest*,
14. Taman Laut / *Sea Park*,
15. Taman Buru, Kebun Buru dan Areal Buru / *Buru Park, Buru Garden and Buru Area*,
16. Wisata Petualangan Alam / *Nature Adventure Tourist*,
17. Wisata Agro / *Agro Tourism*,
18. Kolam Pemancingan / *Fishing Ponds*,
19. Daya Tarik Wisata Buatan/Binaan Manusia Lainnya / *Artificial Tourist Attraction*,
20. Kawasan Pariwisata / *Tourism area*,
21. Aktivitas Taman Bertema atau Taman Hiburan / *Amusement Parks*,
22. Taman Rekreasi / Taman Wisata / *Recreational Park*,
23. Arung Jeram / *Rafting*,
24. Wisata Selam / *Dive Tourism*,
25. Dermaga Marina / *Marina Dock*,
26. Wisata Tirta Lainnya / *Other Water Tourism*.

2.1.6 **Mobile Positioning Data (MPD)**

Mobile Positioning Data (MPD) merupakan suatu dataset berskala besar yang memuat catatan transaksi dan histori lokasi dari pengguna suatu operator seluler (*Mobile Network Operator/ MNO*) yang diproses dan disimpan di dalam suatu sistem (Badan Pusat Statistik, 2022). Berdasarkan jenis pengumpulan data, MPD dapat dibedakan menjadi MPD aktif dan MPD pasif (Ahas, et al., 2007).

MPD aktif merupakan data lokasi ponsel yang didapatkan melalui gelombang sinyal GPS atau sejenisnya. Sedangkan MPD pasif merupakan data lokasi yang tersimpan secara otomatis ke dalam basis data operator seluler ketika ponsel berinteraksi dengan jaringan seluler. MPD pasif terdiri dari *Call Detail Record* (CDR) dan *Location Based Signalling* (LBS). *Call Detail Record* (CDR) mengumpulkan data pengguna operator seluler melalui aktivitas seperti pengiriman dan penerimaan pesan singkat dan aktivitas panggilan. Sedangkan *Location Based Signalling* (LBS) mengumpulkan data pengguna operator seluler yang menggunakan akses internet, updating lokasi wilayah, pencarian jaringan, atau aktivitas lainnya. Aktivitas pesan singkat dan panggilan dalam CDR juga termasuk di dalam LBS ini.

Mobile Positioning Data (MPD) diperoleh dari ponsel yang memanfaatkan jaringan operator seluler dalam penggunaannya. Untuk memperoleh MPD, operator seluler mempunyai cara kerja yang dapat dijelaskan sebagai berikut (Badan Pusat Statistik, 2021) :

1. Sinyal dipancarkan oleh *Base Transceiver Station* (BTS) yang sudah tertanam kode sinkronisasi khusus.
2. *Handphone* akan menerima lokasi identitas dan frekuensi sinyal yang dipancarkan oleh BTS.
3. Setelah menerima sinyal, perangkat akan mengirimkan log ke server, dan server akan melakukan komputasi waktu dan lokasi terjadinya transaksi ini.
4. Data balikan sesuai jenis transaksi kemudian dikembalikan ke ponsel.

Berdasarkan penjelasan diatas, MPD sangat berpotensial untuk menjadi sumber data mobilitas karena MPD memuat data histori lokasi dan waktu. Menurut Ahas et al. (2010), penggunaan MPD memiliki beberapa keunggulan, diantaranya adalah sebagai berikut (Badan Pusat Statistik, 2021) :



1. Penggunaan telepon genggam tersebar luas dan populer di negara maju dan berkembang;
2. Kecenderungan masyarakat untuk selalu membawa telepon genggam dan menjadikannya barang penting;
3. Data awal sudah berbentuk digital sehingga bebas dari *human error* seperti keterbatasan ingatan responden atau kesalahan entri data; dan
4. Penggunaan MPD memungkinkan dalam mempelajari perpindahan penduduk dalam dimensi ruang dan waktu secara lebih tepat.

Meskipun memiliki berbagai potensi dan keunggulan, MPD juga memiliki beberapa keterbatasan yang harus menjadi perhatian, diantaranya (Badan Pusat Statistik, 2021) :

1. Kolaborasi yang rumit dengan MNO sebagai penyedia MPD terkait perlindungan dan kerahasiaan data (Ahas et al, 2011).
2. Isu terkait privasi, akses data, dan pengambilan sampel dari pelanggan MNO (Järv et al., 2012).
3. Biaya eksperimen data yang relatif tinggi serta ukuran data yang besar mengakibatkan rumitnya penanganan dan pemanfaatannya. Kesalahan juga sulit dideteksi dan dikoreksi dalam basis data yang besar (Ahas et al., 2007; 2011).
4. MPD memiliki keakuratan spasial yang cukup rendah (Ahas et al., 2010). Penentuan posisi dan kualitas data MPD dapat sangat berbeda karena perbedaan perangkat keras dan perangkat lunak yang digunakan dalam jaringan seluler.
5. Informasi mengenai karakteristik pengguna tidak dapat diperoleh jika hanya menggunakan MPD tanpa melakukan survei (Ahas et al., 2009).

2.1.7 *Usual Environment*

Usual Environment terdiri dari lingkungan sekitar rumah, tempat bekerja atau belajar, serta tempat-tempat lain yang sering dikunjungi. Konsep *usual environment* memiliki dua dimensi, yakni *vicinity* dan *frequency*. Tempat-tempat

yang terletak dekat dengan tempat tinggal seseorang merupakan bagian dari *usual environment* orang tersebut meskipun jarang dikunjungi. Tempat-tempat yang secara rutin dikunjungi, rata-rata seminggu sekali atau lebih, adalah termasuk *usual environment*, meskipun terletak pada jarak yang cukup jauh dari tempat tinggal. Orang yang bekerja di sarana transportasi seperti pilot, pengemudi truk, dan pramugari dapat dikecualikan oleh kriteria *usual environment* (Cuatomsportal, 2019). IRTS merekomendasikan penentuan *usual environment* harus didasarkan pada kriteria-kriteria berikut:

1. Frekuensi perjalanan
2. Durasi perjalanan
3. Melintasi batas administratif atau nasional
4. Jarak dari tempat tinggal biasa

Menurut IRTS, selain menggunakan kriteria frekuensi dan durasi untuk menentukan lingkungan biasa, disarankan agar penetapan batas *usual environment* menggabungkan praktik lintas batas administrasi dan kriteria jarak untuk menetapkan batas lingkungan biasa. Hal tersebut didasari dengan beberapa alasan, yakni:

1. Unit administratif mungkin memiliki ukuran yang sangat berbeda bahkan di dalam suatu negara
2. Wilayah metropolitan dapat membentang di atas batas administratif meskipun mewakili wilayah geografis yang berdekatan
3. Tempat tinggal biasa beberapa individu mungkin sangat dekat dengan batas administrasi sehingga lalu lintas mereka mungkin tidak relevan untuk analisis pariwisata.

BPS mendefinisikan *usual environment* sebagai wilayah atau area lokasi lingkungan keseharian pengguna seluler yang terdiri dari wilayah rumah, lokasi bekerja, serta rumah kedua pengguna seluler. *Usual environment* untuk setiap pelanggan diidentifikasi pada setiap bulan. Karakteristik *home* dan *work* teridentifikasi dengan memperhitungkan dimensi geografis dan temporal selama

dua belas bulan ke belakang. Mekanisme tersebut dibangun dengan tujuan agar MPD dapat menyesuaikan dengan konsep domisili penduduk. Selain itu, dengan melihat riwayat yang lebih panjang, akan dihasilkan identifikasi lokasi *usual environment* yang lebih akurat dan tidak terpengaruh oleh pergerakan sementara pelanggan di luar *usual environment* asli mereka, seperti berlibur, bekerja di luar kota, dan lain-lain (Cuatomsportal, 2019).

2.1.8 AMDA

Anchor Mobility Data Analytic (AMDA) merupakan suatu algoritma yang digunakan untuk menentukan *Homework Location* dari *Mobile Positioning Data* (A. P. Putra et al, 2021). Algoritma ini memanfaatkan pasif *Mobile Positioning Data* berupa *Call Detail Record* (CDR) dalam menentukan *Homework Location*. CDR adalah log yang dibuat secara otomatis yang berisi penggunaan ponsel aktif dalam jaringan, seperti panggilan masuk dan keluar, SMS, GPRS, dll. Data transaksi yang direkam juga disertai dengan informasi waktu aktivitas dan identitas pelanggan berupa nomor *handphone*. Identifikasi lokasi ditetapkan dari rekaman lokasi BTS penyedia jaringan seluler dalam bentuk koordinat posisi (A. P. Putra et al, 2021).

Algoritma AMDA dilakukan dengan melalui beberapa tahap. Tahap pertama adalah pembentukan *anchor model*. Pembentukan *anchor model* adalah proses penghapusan data transfer lokasi berkecepatan tinggi dari data mentah ke data master dengan menghitung jarak urutan setiap lokasi MSISDN dengan menerapkan jarak Haversine dan menghitung kecepatan antar lokasi (A. P. Putra et al, 2021). Setelah menentukan *anchor*, langkah selanjutnya adalah menentukan *Usual Environment Candidate* dengan tahapan sebagai berikut:

1. Menghitung jarak antara *anchor* dan *main candidate*
2. Melakukan filter *anchor* dengan jumlah hari lebih dari atau sama dengan 5.
3. Hilangkan *anchor* yang dekat dengan *main candidate* (di bawah 500 m).
4. Memisahkan waktu hanya pada hari kerja, serta transformasi variabel waktu menjadi format 12 jam.

4. Melakukan perhitungan frekuensi lokasi awal harian setiap titik lokasi (*anchor*) diikuti dengan statistik rata-rata dan standar deviasi setiap jam kejadian, dan dijadikan *main candidate*.
5. Menghitung jarak antara lokasi terpilih dengan calon lokasi utama kemudian menghilangkan lokasi berdekatan yang berada di bawah 500 m, dilanjutkan dengan pengurutan hari kejadian di atas atau sama dengan 5 hari (A. P. Putra et al, 2021).

Setelah melalui tahapan-tahapan di atas, Algoritma AMDA akan menghasilkan prediksi *usual environment*. Titik *anchor* dengan rata-rata jam kurang dari tujuh ($\text{AVG_hour} < 7$) akan diklasifikasikan sebagai kandidat lokasi rumah, sedangkan *anchor* lain dengan rata-rata jam lebih dari tujuh akan diklasifikasikan sebagai calon lokasi kerja (A. P. Putra et al, 2021).

2.1.9 *Family Grouping*

Family atau keluarga didefinisikan sebagai kelompok yang terdiri dari dua orang atau lebih yang tinggal bersama dan memiliki hubungan berdasarkan kelahiran, perkawinan, ataupun adopsi (Tillman, K. H. et al, 2008). Secara umum, penentuan *family grouping* didasarkan pada prinsip kesamaan karakteristik tertentu dan kombinasi yang identik menggunakan sistem klasifikasi (Sedqui, A. et al, 1995). Dasar pengelompokan individu yang termasuk dalam satu rumah tangga dapat diidentifikasi dari *usual environment* dan pola perjalanan wisata.

Sumber *big data* secara umum terbagi dalam tiga kategori, yaitu *User Generated Content* (UGC) yang mencakup data sosial media berupa teks maupun foto, *device data* yang berisi informasi mengenai telepon dan GPS, serta data transaksi. *Usual environment* dan pola perjalanan wisata seseorang dapat dideteksi menggunakan *device data* dan data transaksi individu dalam kesehariannya. Metode *family grouping* dapat dilakukan dengan memanfaatkan algoritma *machine learning* yang diimplementasikan pada *big data* (Derdouri, A dan Osaragi, T, 2021).

2.1.10 *Double Counting*

Double Counting adalah penghitungan orang dengan kepemilikan *sim card* yang dijangkau oleh penyedia layanan (*provider*) lebih dari satu kali dalam periode pelaporan yang sama. Penghitungan ganda harus dihindari karena dapat meningkatkan jumlah orang yang dicapai (Chaplowe, S. G, 2017). Dalam konsep *Mobile Positioning Data* (MPD), *double counting* terjadi ketika individu yang sama memiliki nomor telepon yang berbeda dalam satu perangkat. Seharusnya individu tersebut terdeteksi lebih dari satu kali pada lokasi dan waktu yang sama. Namun, dikarenakan jarak tower seluler (BTS) tiap operator dan aktivitas seluler setiap pengguna berbeda, individu tersebut mempunyai *record* MPD yang berbeda. *Double counting* akan mengakibatkan *overestimate* pada sebuah estimator, sehingga estimator tersebut tidak layak digunakan sebagai penaksir populasi (Singh, J, 2018). Penanganan *double counting* dalam *Mobile Positioning Data* dapat dilakukan dengan membangun algoritma yang sesuai.

2.1.11 *Machine Learning*

Machine learning merupakan salah satu bentuk *artificial intelligence* di mana algoritma atau program komputer dapat menemukan dan mengaplikasikan pola-pola dalam data. Algoritma *machine learning* dibangun untuk menyelesaikan tiga kebutuhan, yaitu prediksi, klasifikasi, dan *clustering* (Kusuma, 2020). Terdapat empat jenis *machine learning* berdasarkan teknik pembelajarannya, yaitu *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning*.

Supervised learning merupakan teknik *machine learning* yang menggunakan *data training* yang sudah berlabel. Data tersebut diidentifikasi sebagai label input yang selanjutnya digunakan untuk melakukan prediksi dan klasifikasi. Algoritma yang termasuk dalam teknik *supervised learning* adalah Decision Tree, K-Nearest Neighbor (KNN), Naive Bayes, Regresi, dan Support Vector Machine (Retnoningsih, Endang dan Pramudita, 2020). Berbeda dengan *supervised learning*, teknik *unsupervised learning* akan mengembangkan klasifikasi label secara otomatis. Algoritma *unsupervised learning* bertujuan

untuk mencari kesamaan karakteristik dari objek dalam data yang kemudian dikelompokkan menjadi sebuah *cluster* secara otomatis. Algoritma yang termasuk dalam *unsupervised learning* diantaranya k-means, apriori, Independent Subspace Analysis (ASI), dan DBSCAN (Ahmad et al., 2019). Sementara itu, *semi-supervised learning* berada di antara *supervised learning* dan *unsupervised learning*. *Semi-supervised learning* menggabungkan data berlabel dengan data yang tidak berlabel selama training (Pratama, 2020).

2.1.12 Python

Python merupakan salah satu bahasa pemrograman yang dapat digunakan untuk berbagai macam keperluan, seperti pembuatan aplikasi hingga melakukan analisis data (Vallat, 2018). *Python* pertama kali dicetuskan di Belanda oleh Guido van Rossum pada tahun 1991. Sintaks *Python* didasarkan pada kombinasi beberapa bahasa pemrograman yang ada dengan pengembangan dominan berasal dari bahasa pemrograman ABC (Awangga, et al., 2019). Saat ini *Python* sudah dapat dioperasikan di berbagai platform sistem operasi, seperti *Windows*, *Mac OS*, *Android*, dan *Linux*. *Python* bersifat *open source* dan berlisensi gratis sehingga mudah digunakan dan didistribusikan.

Python sangat cocok untuk pemula yang ingin mempelajari ilmu data, pengembangan situs, dan *machine learning* karena *Python* memiliki kode pemrograman yang sederhana, lengkap, dan mudah dipahami (Vallat, 2018). Keterbacaan serta pemahaman *Python* berasal dari penggunaan sistem indentasi, yaitu pemindahan antara blok-blok program susunan indentasi. *Python* memiliki istilah lain seperti bahasa pemrograman interpreter. Menurut Kurniawan, et al. (2005), secara *default* interpreter bawaan *Python* adalah *Integrated Development and Learning Environment* (IDLE). Interpreter digunakan untuk dua jenis: mode baris perintah dan mode skrip. Di bawah ini adalah kelebihan *Python* dibandingkan bahasa pemrograman lain:

- Efisiensi tinggi
- Efektif
- Dapat dikompilasi dengan bahasa pemrograman lain

- Banyak modul tersedia, dan
- Berorientasi multi-paradigma (objek dan prosedur secara bersamaan).

2.1.13 *Haversine Distance*

Haversine merupakan sebuah metode untuk menghitung jarak antara dua titik dengan memperhitungkan bahwa bumi bukanlah sebuah bidang datar namun adalah sebuah bidang yang memiliki derajat kelengkungan (Putra et al., 2019). Teorema *Haversine* digunakan untuk menghitung jarak antara 2 titik pada permukaan bumi berdasarkan garis bujur (*latitude*) dan garis lintang (*longitude*). Rumus Haversine mengasumsikan pengabaian efek *ellipsoidal*, cukup akurat untuk sebagian besar perhitungan, juga pengabaian ketinggian bukit dan kedalaman lembah di permukaan bumi (Putra et al., 2019). Dalam formula haversine, jarak antara dua titik pada permukaan bumi (yang diasumsikan berbentuk bulat sempurna) dihitung dengan menggunakan persamaan berikut:

$$\begin{aligned} d &= 2r \arcsin(\sqrt{\text{hav}(\varphi_2 - \varphi_1) + (1 - \text{hav}(\varphi_1 - \varphi_2) - \text{hav}(\varphi_1 + \varphi_2)).\text{hav}(\lambda_2 - \lambda_1)}) \\ &= 2r \arcsin(\sqrt{\sin^2(\frac{\varphi_2-\varphi_1}{2}) + (1 - \sin^2(\frac{\varphi_2-\varphi_1}{2}) - \sin^2(\frac{\varphi_2+\varphi_1}{2})).\sin^2(\frac{\lambda_2-\lambda_1}{2})}) \\ &= 2r \arcsin(\sqrt{\sin^2(\frac{\varphi_2-\varphi_1}{2}) + \cos \varphi_1.\cos \varphi_2.\sin^2(\frac{\lambda_2-\lambda_1}{2})}) \end{aligned}$$

dengan :

d adalah jarak antara dua titik pada permukaan bumi

φ_1, φ_2 adalah koordinat latitude (dalam format radians) dari titik 1 dan titik 2,

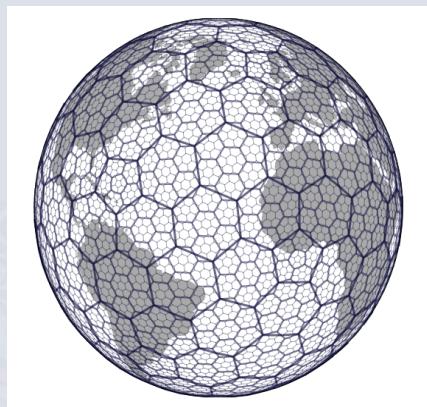
λ_1, λ_2 adalah koordinat longitude (dalam format radians) dari titik 1 dan titik 2.

2.1.14 *Hexagonal Hierarchical Geospatial Indexing System (H3)*

Hexagonal Hierarchical Geospatial Indexing System (H3) merupakan sistem pengindeksan *geospasial* untuk mengelompokkan data spasial berdasarkan *grid* yang berbentuk heksagonal atau segi enam. Gagasan utamanya adalah segi enam (heksagon) merupakan bentuk yang paling bermanfaat untuk analisis spasial dan dapat digunakan untuk membuat perhitungan yang konsisten (Jordan Bean,

2022). Hal tersebut dikarenakan bentuk heksagonal membuat semua tetangga memiliki jarak yang sama (Uber Technologies, Inc., 2023).

Visualisasi Skema Pengindeksan Geospasial dengan H3



Gambar 2.1 Sistem Pengindeksan Geospasial dengan H3

Sumber : Isaac Brodsky, 2018

Setiap heksagon pada H3 memiliki nomor identitas unik yang dikenal sebagai H3 *index*. H3 index dapat digunakan untuk memetakan lokasi *geospasial* pada permukaan bumi ke dalam sebuah heksagon dan juga untuk menghitung jarak antara dua lokasi yang diwakili oleh dua heksagon yang berbeda. Sistem H3 memiliki struktur hierarkis yang memungkinkan identifikasi unik dan hubungan antara sel-sel heksagonal pada berbagai tingkat resolusi. H3 memiliki 15 resolusi dengan rincian luas area tiap resolusi adalah sebagai berikut.

Tabel 2.1 Resolusi H3

Resolution	Average Hexagon Area (km ²)
0	4,357,449.42
1	609,788.44
2	86,801.78
3	12,393.43
4	1,770.35



5	252.9038582
6	36.12906216
7	5.16129336
8	0.737327598
9	0.105332513
10	0.015047502
11	0.002149643
12	0.000307092
13	0.00004387
14	0.000006267
15	0.000000895

Sumber : Uber Technologies, Inc., 2023

Adanya tingkat resolusi tersebut memungkinkan indeks dan pengambilan data *geospasial* yang efisien serta operasi spasial seperti agregasi dan filtrasi. Dengan menggunakan H3, pengguna dapat dengan mudah mengkonversi koordinat lintang dan bujur menjadi indeks H3, menemukan sel-sel tetangga, dan melakukan perhitungan spasial berdasarkan struktur grid heksagonal (Uber Technologies, Inc., 2023). Menurut Andrey Sekste & Eduard Kazakov (2022), purwarupa H3 yang berbentuk heksagonal memiliki beberapa kelebihan, diantaranya adalah:

1. Jarak yang sama antar *centroid* heksagon yang berdekatan dari segala arah membuat penghitungan jarak ke daerah tetangga menjadi lebih mudah dan lebih efisien. Hubungan spasial heksagon yang identik dengan sel di sekitarnya menyederhanakan pemrosesan *query* spasial.
2. Heksagon memiliki rasio *perimeter-to-area* yang rendah atau bentuk poligonal paling "melingkar" yang mengurangi bias pengambilan sampel

yang diciptakan oleh efek tepi. Terdapat lebih sedikit *outlier* daripada di kotak atau segitiga dengan ukuran yang sama.

3. Geometri segi enam memudahkan representasi kurva objek kehidupan nyata seperti jalan dan sungai.
4. Di area yang luas, kelengkungan bumi menyebabkan lebih sedikit distorsi pada kisi heksagonal daripada kisi persegi.

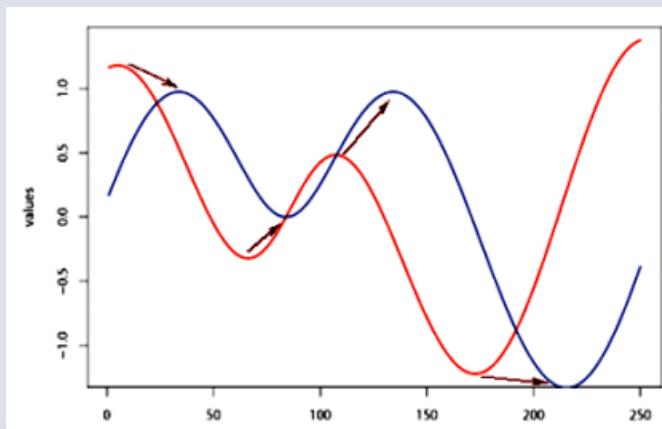
Selain kelebihan, purwarupa H3 juga memiliki beberapa kekurangan, diantaranya:

1. Jaringan heksagonal tidak selalu optimal secara universal. Bahkan, seringkali ditemukan jaringan kotak di dalam analisis GIS dan pemetaan tematik. *H3 Hexagons* adalah suatu sistem koordinat yang menggambarkan bumi dalam bentuk heksagonal. Namun, ketika suatu heksagon H3 mencakup wilayah dengan batas-batas tertentu seperti air dan permukaan tanah, dapat menyebabkan masalah dalam mewakili data yang tepat di dalam heksagon tersebut. Misalnya dalam penghitungan kepadatan penduduk.
2. H3 memiliki kesulitan tersendiri untuk disampaikan pada *audiens*.
3. Menggunakan H3 akan membuat keterangan terkait wilayah administratif hilang. Dalam banyak kasus, alasan yang menentukan batasan administratif mungkin tidak relevan dengan kasus penggunaan. *Centroid* bisa saja digunakan untuk menentukan wilayah administratif dari suatu heksagon, tapi bisa jadi tidak relevan

2.1.15 *Dynamic Time Warping*

Dynamic Time Warping (DTW) adalah metode untuk menghitung kesamaan antara dua data deret waktu (*time series*) yang dapat berbeda dalam waktu dan kecepatan (BlazStrle et al, 2003). Ilustrasi mengenai DTW dapat dilihat pada gambar 2.2.

Ilustrasi DTW



Gambar 2.2 Ilustrasi DTW

Sumber: Gunawan & Salim, 2013

Pada gambar di atas, terdapat dua grafik yang menunjukkan dua data deret waktu yang memiliki kesamaan pola, namun dengan letak posisi awal yang berbeda. Hal ini terjadi karena adanya perbedaan selang waktu antara X dan Y. Permasalahannya adalah bagaimana cara mengidentifikasi bahwa kedua data deret waktu tersebut dapat dianggap mempunyai pola yang sama walaupun memiliki dua rentang waktu dan perubahan kecepatan yang berbeda. Dengan metode *Dynamic Time Warping*, masalah ini dapat dipecahkan dengan membandingkan biaya *optimal path warping* dengan nilai ambang batas yang ditentukan. Berikut algoritma dan kriteria yang diuraikan oleh (Senin, P, 2008) untuk menghitung *optimal path warping*:

1. Menentukan Matriks *Cost*

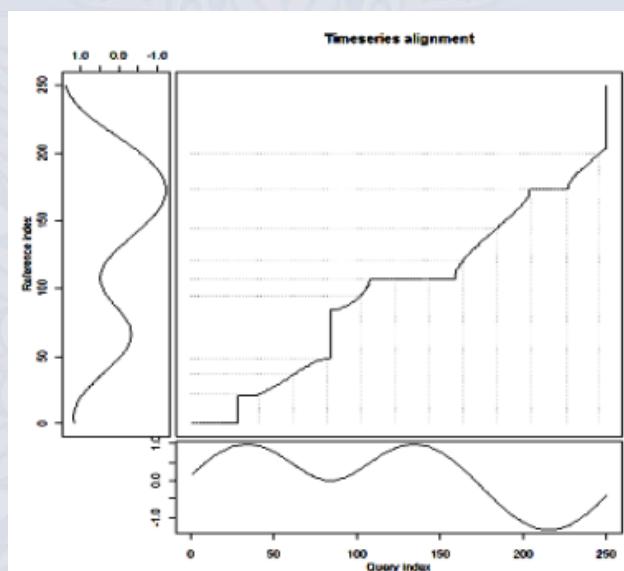
Matriks *cost* adalah matriks yang merepresentasikan selisih dari himpunan X dan Y. Kemudian nilai dari *cost* setiap matriksnya digunakan untuk mencari jalur dengan *cost* terendah dari kolom pertama sampai kolom terakhir. Masing-masing elemen matriks dapat dihitung dengan persamaan berikut:

$$C_p = [C_{1,1} \ C_{1,2} \dots \ C_{1,f} \ C_{2,1} \ C_{2,2} \dots \ C_{2,f} : C_{f,1} : \vdots : C_{f,2} \dots : C_{1,f}]$$

2. Menentukan Optimal Warping Path

Setelah matriks $cost$ terbentuk, algoritma ini akan mencari jalur $warping$ dengan $cost$ terendah $p = (p_1, p_2, \dots, p_k)$ dengan $p_i = (p_i, p_j) \in [1: n] \times [1: m]$ dan $I \subset [1: k]$. Optimal $warping path$ seperti pada gambar 2.3 adalah jalur dengan biaya ($cost$) terpendek dari selisih perbandingan X dan Y.

Optimal Warping Path dari Dua Deret Waktu X dan Y



Gambar 2.3 Optimal Warping Path dari 2 Deret Waktu X dan Y

Sumber: Gunawan & Salim, 2013

3. Menentukan Matriks DTW

Berdasarkan definisi dari optimal $warping path$ dalam pencarian jalur dengan matriks $cost$, diperlukan pengujian ke setiap jalur antara X dan Y. Hal ini tentu saja tidak efisien karena kemungkinan jalur yang berpotensi untuk dilewati juga akan semakin banyak seiring dengan panjang deret waktu X dan Y. Untuk mengatasi masalah ini, digunakan *dynamic programming* sebagai metode optimasi pencarian jalur dengan cepat. Menurut (Levitin, A, 2006) *dynamic programming* adalah sebuah teknik untuk menyelesaikan masalah dengan cara membagi masalah dalam beberapa sub masalah (*stage*) sehingga

solusi dari masalah dapat dipandang sebagai serangkaian keputusan yang saling berkaitan. Dalam kasus ini dilakukan perhitungan untuk total *cost* setiap pencarian jalur yang mungkin dilewati (*stage*). Semua jalur yang telah dilewati diakumulasi terlebih dahulu. Matriks akumulasi yang akan digunakan didefinisikan sebagai berikut:

$$DTW(X, Y) = C_p(X, Y)$$

4. Menentukan *Cost Optimal Warping Path*

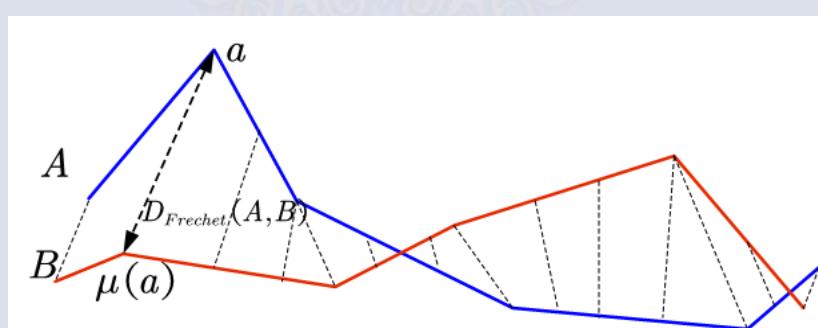
Setelah menghitung matriks DTW, penentuan total dari *cost* jarak yang melewati jalur *warping* dapat menggunakan persamaan di bawah ini dengan *cost* dari optimal *warping path* adalah *cost* akumulasi minimum pada kolom terakhir.

$$\text{cost} = \frac{1}{m} \{ \min[DTW(1, m), DTW(2, m), \dots, DTW(n, m)] \}$$

2.1.16 *Frechet Distance*

Frechet Distance adalah sebuah metode dalam analisis jarak antara dua kurva dalam ruang metrik untuk melihat kesamaan antara kedua kurva tersebut (Alt, H., & Godau, M., 1995). Konsep dasar dari *Frechet Distance* adalah menemukan jarak terpendek di antara semua kemungkinan jalan atau lintasan antara dua kurva atau lintasan tersebut. Ilustrasi mengenai metode ini dapat terlihat pada gambar 2.4.

Ilustrasi Frechet Distance



Gambar 2.4. Ilustrasi Frechet Distance

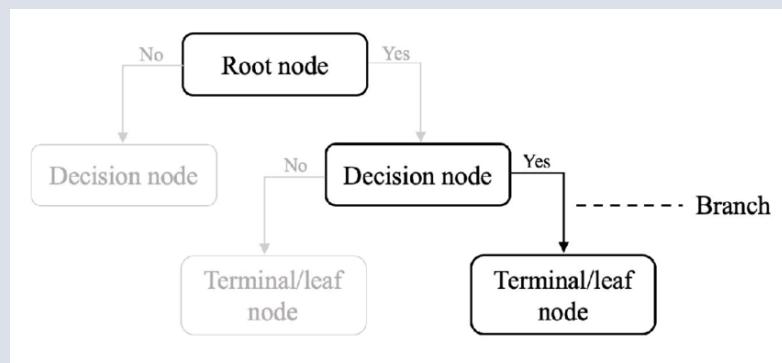
Sumber : Guo et al., 2017

Menurut Alt & Godau (1995), cara kerja *Frechet Distance* didasarkan pada prinsip perjalanan dua agen, dimana setiap agen bergerak sepanjang kurva atau jalur/lintasan tertentu dengan kecepatan yang konstan. Satu agen bergerak di sepanjang kurva pertama, sementara agen lainnya bergerak di sepanjang kurva kedua. Jarak *Frechet* adalah panjang jalur terpendek yang mungkin untuk menghubungkan kedua agen ini, dengan asumsi bahwa keduanya tidak pernah bergerak mundur. Struktur simpul-simpul di dalam lintasan dipertimbangkan dalam proses perhitungan *Frechet Distance* sehingga secara lebih akurat dapat menggambarkan kesamaan antar lintasan-lintasan tersebut. Sebagai contoh, ketika terdapat arah mundur, memutar, atau silang dalam suatu lintasan, nilai *Fréchet Distance* tidak mengalami distorsi seperti pengukuran jarak lainnya. Oleh karena itu, metrik *Frechet Distance* lebih deskriptif dan cocok sebagai ukuran kemiripan antar lintasan atau kurva (Guo, et al., 2017).

2.1.17 *Decision Tree* dan *Gini Impurity*

Menurut Khazari et al. (2017), *decision tree* (pohon keputusan) merupakan salah satu metode yang ada pada teknik klasifikasi dalam *data mining*. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record* dimana atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon (Khazari et al., 2017). Pohon keputusan merupakan struktur pohon yang terdiri dari simpul-simpul (*nodes*) yang mewakili keputusan atau prediksi dengan fitur/atribut tertentu, serta cabang-cabang (*branches*) yang menghubungkan simpul-simpul tersebut dan mewakili kemungkinan nilai atau kondisi yang dapat diambil oleh fitur tersebut. Adapun setiap simpul daun (*leaf node*) pada pohon keputusan menunjukkan hasil atau prediksi akhir (Lendave, 2021).

Struktur *Decision Tree*



Gambar 2.5 *Decision Tree*

Sumber : Lendave, 2021

Ide utama di balik pohon keputusan adalah mencari fitur-fitur yang menyimpan informasi paling banyak tentang fitur target, lalu mempartisi dataset berdasarkan nilai dari fitur tersebut sehingga dihasilkan nilai-nilai fitur target yang semurni mungkin pada setiap simpul. Fitur yang paling informatif adalah yang paling baik dalam mengisolasi ketidakpastian informasi tentang fitur target. Proses pencarian untuk fitur yang paling informatif berlangsung hingga berakhir dengan simpul daun yang murni (Lendave, 2021).

Tujuan utama dari pohon keputusan adalah menemukan pemisah terbaik (*best split*) antara simpul-simpul yang akan membagi data ke dalam kategori yang sesuai. Salah satu metode yang dapat digunakan untuk menemukan simpul sebagai pemisah terbaik adalah *gini impurity* (Lendave, 2021). *Gini impurity* adalah jenis ukuran yang dapat digunakan untuk mengukur kesalahan dalam pelabelan yang tidak sesuai dengan pola data dan dapat digunakan dalam pembangunan pohon keputusan sebagai alternatif untuk entropi (Suthaharan, 2016). Derajat *gini impurity* berkisar dari 0 hingga 1, dengan 0 menunjukkan bahwa semua elemen termasuk dalam satu kelas dan 1 menunjukkan bahwa hanya ada satu kelas. *Gini impurity* dengan nilai 1 menunjukkan bahwa semua item tersebar secara acak di berbagai kelas, sedangkan nilai 0,5 menunjukkan bahwa elemen-elemen tersebar merata di beberapa kelas (Lendave, 2021).

Zhi et al. (2018) menjelaskan konsep *gini impurity* adalah ketika diberikan satu set D berisi *records* dari n kelas sedemikian sehingga p_j merepresentasikan frekuensi relatif kelas j di dalam D . *Gini Impurity* ($\text{Gini}(D)$) kemudian didefinisikan sebagai:

$$\text{Gini}(D) = 1 - \sum_{j=1}^n p_j^2$$

Gini impurity mencerminkan tingkat ketidakmurnian dari sekumpulan informasi. Ketika *record* didistribusikan secara merata pada semua kelas, $\text{Gini}(D)$ akan mencapai nilai maksimum ($1 - 1/N$), dan akan mencapai nilai minimum nol ketika semua *record* hanya terdistribusi dalam satu kelas.

2.2 Penelitian Terkait

Penggunaan *Mobile Positioning Data* (MPD) yang dipadukan dengan survei digital (*combination methods*) untuk pemanfaatan penelitian sudah diterapkan oleh BPS. Penelitian terkait penggunaan MPD dalam mendukung penggunaan *big data* dalam *official statistics* pernah dilakukan oleh Putra, et al (2021). Hasil penelitian tersebut menunjukkan bahwa *combine methods* memiliki keuntungan yaitu estimasi dapat dilakukan di wilayah yang lebih kecil (kabupaten/kota), periode data yang diperoleh dapat berupa data bulanan, tidak memerlukan petugas pendataan, dapat digunakan untuk perhitungan statistik lainnya seperti komuter dan migrasi, serta biaya yang dikeluarkan hanya setengah dari biaya dari survei konvensional.

Pada penelitian ini, penggunaan MPD dipadukan dengan keunggulan survei untuk mendapatkan data yang lebih lengkap. Survei digital dilakukan untuk melengkapi CKG untuk mendapatkan variabel penting lainnya seperti demografi, tujuan perjalanan, mode transportasi, akomodasi, jenis pengeluaran, dan jumlah pengeluaran. Dalam pemodelan data, BPS menggunakan data dari operator telepon seluler yang telah dibuat anonim sehingga BPS tidak mengetahui siapa yang bepergian. Namun, untuk mengevaluasi akurasi algoritma, BPS merekrut relawan yang bersedia menyerahkan log data selulernya ke BPS sebagai data



sampel untuk kemudian dianalisis dan diverifikasi perjalanannya. Akurasi prediksi lokasi rumah di tingkat kabupaten/kota terus meningkat seiring dengan perbaikan algoritma, dengan nilai akurasi mencapai 96,23% pada tahun 2018.

Penelitian lain terkait penggunaan *Mobile Positioning Data* (MPD) juga pernah dilakukan oleh Munaf et. al (2021). Penelitian ini bertujuan untuk menjelaskan matriks kualitas yang digunakan oleh BPS dalam memeriksa data dari operator seluler untuk penjaminan kualitas data MPD. BPS menggunakan tiga belas indikator dalam melakukan penjaminan mutu. Pemeriksaan tersebut menggunakan beberapa metode yang berbeda, seperti penetapan ambang batas, pemeriksaan kelengkapan data, dan pemeriksaan formulir distribusi data. Periode data yang digunakan dalam penelitian ini adalah data operator jaringan seluler pada bulan Juni sampai Juli 2020 sebagai dasar analisis MPD tahun 2021. Penelitian tersebut menyatakan bahwa BPS dapat bekerja sama dengan operator seluler terkait untuk melakukan analisis data pada tahun 2021. Namun, operator harus mengulang perhitungan matriks yang dibutuhkan sebagai penjaminan mutu setiap bulan untuk memastikan kondisi data tetap terjaga dan sesuai dengan standar yang telah ditetapkan sebelumnya. BPS memiliki rencana untuk berkolaborasi dengan MNO lain dan perlu menyesuaikan indikator serta ambang batasnya berdasarkan jenis data yang dimiliki MNO.

Beberapa studi telah mengembangkan berbagai pendekatan untuk mendeteksi *usual environment*, terutama lokasi rumah (*home*) and kantor (*work*). Sebagian besar algoritma pendekripsi *home* dan *work location* yang diusulkan menggunakan *Call Detail Record* (CDR). Namun, CDR memiliki keterbatasan sebagai sumber informasi lokasi karena CDR dibuat hanya ketika ada panggilan suara atau pesan teks. Pada saat ini, sebagian besar pengguna seluler lebih banyak menggunakan internet, bahkan saat melakukan panggilan sehingga jumlah rekaman yang dihasilkan CDR tidak signifikan dibandingkan dengan aktivitas seluler lainnya. Putra, et al (2021) mengusulkan sumber data pemosisian seluler lain seperti *Location-Based Advertising* (LBA) untuk menyimpulkan *home* dan *work location*. Oleh karena itu, *Anchor Mobility Data Analytic* (AMDA)

diusulkan untuk menentukan *home* dan *work location* dari data pemosisian seluler. Algoritma ini menggunakan pembalikan searah jarum jam untuk memudahkan mengklasifikasikan seseorang di lingkungan biasanya. Penelitian ini menyimpulkan bahwa ketepatan AMDA dalam menentukan lokasi rumah bulanan adalah 98,8% di tingkat provinsi dan 88,7% di tingkat kabupaten. Sedangkan untuk penetapan lokasi kerja bulanan adalah 98,9% di tingkat provinsi dan 70,4% di tingkat kabupaten.

Penelitian yang dilakukan oleh Zhou, et al. (2020) mengusulkan sebuah algoritma indeksasi sel dari area spasial berlingkar berdasarkan struktur *hexagonal lattice quad-tree* (HLQT). Algoritma ini menggunakan operasi kode tambahan dan struktur *quad-tree* untuk mempercepat *indexing* hierarki. Hasil penelitian menunjukkan bahwa efisiensi pengindeksan satu tingkat dari algoritma yang diusulkan sekitar dua kali lipat dari metode tradisional dan efisiensi pengindeksan hierarkis sekitar 67 kali dari metode tradisional. Temuan ini memverifikasi kelayakan dan keunggulan algoritma yang diusulkan dalam penelitian ini.

Selain itu, terdapat penelitian oleh Toohey, et. al. (2015) yang membahas empat ukuran kesamaan lintasan yang paling umum yaitu *longest common subsequence* (LCSS), jarak Frechet, *dynamic time warping* (DTW), dan jarak edit. Keempat ukuran tersebut dapat ditemukan pada paket R yang dinamakan “SimilarityMeasures” dan tersedia di CRAN. Keempat ukuran ini memiliki tujuan yang sama, tetapi memiliki konsep yang berbeda. Jarak Frechet melakukan penghitungan kesamaan lintasan dengan acuan mencari jarak minimum yang diperlukan untuk menyelesaikan lintasan kedua kurva, DTW melakukan pencocokan titik dengan cara nonlinear, LCSS melakukan pencarian jumlah maksimum titik yang dapat dianggap setara saat lintasan dilalui secara monoton dari awal hingga akhir, dan jarak edit menggunakan konsep penghitungan minimum suntingan yang diperlukan untuk membuat dua lintasan menjadi sama. Pada penelitian ini, ukuran-ukuran tersebut akan diperbandingkan menggunakan dataset lintasan pengemudi pengiriman di Inggris dan disoroti perbedaan serta



persamaan yang akan dihasilkannya. Hasil penelitian menunjukkan terjadi korelasi kuat positif antara jarak Frechet dengan DTW dan korelasi kuat negatif antara rasio LCSS dengan rasio jarak edit. Rasio LCSS dan jarak edit memiliki rentang 0 hingga 1 dimana 1 pada LCSS berarti kasus paling mirip sedangkan 1 pada jarak edit berarti kasus sangat berbeda. Oleh karena itu, baik pasangan jarak Frechet dengan DTW dan rasio LCSS dengan rasio jarak edit memiliki kesimpulan yang sama, yaitu keduanya akan menghasilkan hasil yang serupa. Namun, kemungkinan pasangan jarak Frechet dengan DTW memiliki hasil yang berbeda dapat juga terjadi dikarenakan keduanya tidak memiliki batasan ketidaksamaan sehingga sensitif terhadap adanya outlier. Hal ini berbeda dan tidak terjadi dengan pasangan LCSS dan jarak edit karena keduanya memiliki rasio maksimum 1.

2.3 Kerangka Pikir

Penelitian ini dimulai dengan menerapkan penjaminan kualitas (QA) data MPD yang dihasilkan dari nomor-nomor relawan. Hal tersebut dilakukan karena data MPD sendiri memiliki karakteristik *veracity* (kualitas, keakuratan, dan ketidakpastian data) sehingga penjaminan kualitas diperlukan untuk memastikan data yang akan digunakan pada proses selanjutnya memiliki kualitas yang cukup baik dan layak untuk diolah. Adapun indikator-indikator penjaminan kualitas data MPD pada penelitian ini diperoleh melalui studi literatur.

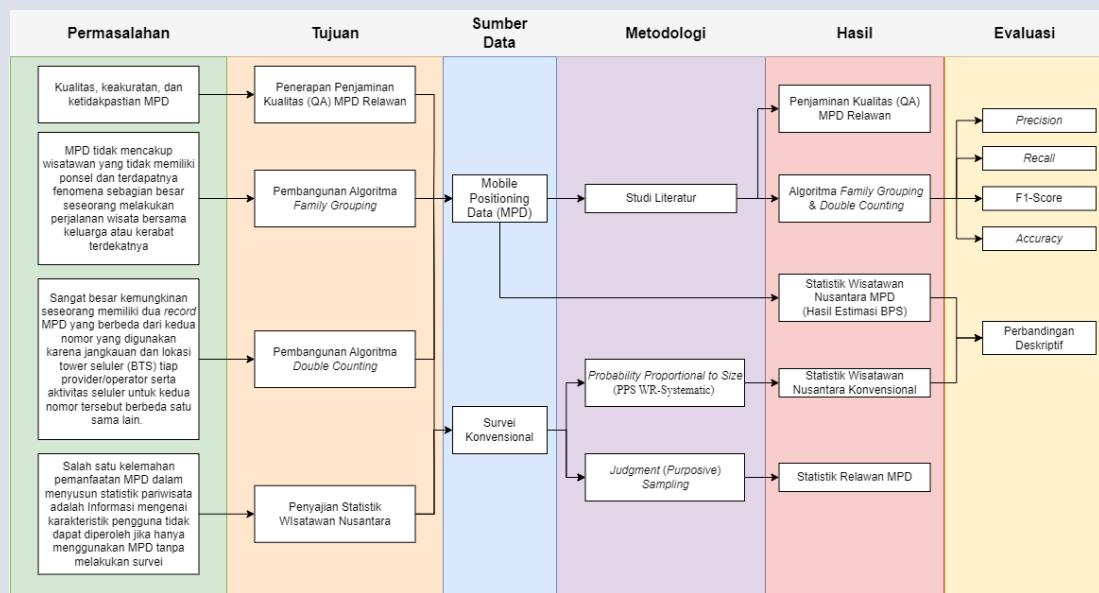
Penelitian ini mencoba untuk memanfaatkan MPD dalam menentukan seseorang termasuk ke dalam satu keluarga (*family grouping*) berdasarkan *usual environment* dan pola perjalanan wisata yang dilakukan. Hal tersebut dilakukan karena adanya permasalahan MPD yang tidak mencakup wisatawan yang tidak memiliki ponsel dan adanya fenomena dimana sebagian besar perjalanan wisata seseorang dilakukan bersama keluarga atau kerabat terdekatnya (Sriyanto & Sidik, 2017) sehingga pendekatan yang dapat dilakukan adalah dengan mengestimasi perjalanan wisata yang dilakukan bersama orang terdekatnya, seperti keluarga. Selain itu, penelitian ini juga akan mencoba untuk mengatasi permasalahan terkait adanya perbedaan *record* MPD pada dua nomor yang digunakan dalam satu



ponsel secara bersamaan yang diakibatkan oleh jangkauan dan lokasi tower seluler (BTS) tiap operator serta aktivitas seluler untuk kedua nomor tersebut berbeda satu sama lain. Upaya mengatasi kedua permasalahan tersebut dilakukan dengan membangun algoritma yang sesuai melalui studi literatur. Algoritma yang dibangun diharapkan dapat mengelompokkan seseorang termasuk ke dalam satu keluarga (*family grouping*) dan mengatasi permasalahan *double counting* dengan baik. Untuk mengukur kinerja algoritma yang telah dibangun, akan dilakukan evaluasi terhadap algoritma melalui beberapa ukuran evaluasi, seperti *precision*, *recall*, *f1-score*, dan akurasi.

Selain itu, karena salah satu kelemahan pemanfaatan MPD dalam menyusun statistik pariwisata ialah informasi mengenai karakteristik pengguna yang tidak dapat diperoleh apabila hanya menggunakan MPD tanpa melakukan survei (Ahas et al., 2009), penelitian ini mencoba untuk menyajikan statistik wisatawan nusantara, terutama seputar perjalanan wisata, menggunakan pendekatan MPD dan survei konvensional. Pendekatan MPD dilakukan dengan memperoleh hasil estimasi statistik wisatawan nusantara berupa perjalanan wisata menggunakan MPD yang didapatkan dari Badan Pusat Statistik (BPS). Di sisi lain, pendekatan survei konvensional dilakukan dengan mengestimasi statistik wisatawan nusantara menggunakan data hasil survei konvensional yang dilakukan pada penelitian ini. Hasil estimasi dengan dua pendekatan berbeda tersebut akan dibandingkan secara deskriptif. Pada penelitian ini juga akan ditampilkan beberapa statistik mengenai karakteristik relawan dalam penelitian ini. Berdasarkan uraian di atas, dibentuklah kerangka pikir yang menjadi acuan dalam penelitian ini, yakni seperti yang terlihat pada gambar 2.6.

Kerangka Pikir



Gambar 2.6 Kerangka Pikir

BAB III

METODOLOGI

3.1 Ruang Lingkup Penelitian

3.1.1 Waktu dan Tempat Penelitian

Praktik Kerja Lapangan (PKL) Politeknik Statistika STIS Riset 1 Tahun Akademik 2022/2023 terdiri dari beberapa rangkaian kegiatan yang dimulai dari kegiatan pra-pencacahan sampai pasca pencacahan. Kegiatan pra-pencacahan meliputi penyusunan topik penelitian dan desain survei yang dimulai pada tanggal 23 Oktober 2022 sampai dengan 15 November 2022. Kegiatan selanjutnya dilakukan survei pendahuluan sebanyak satu kali. Survei Pendahuluan bertujuan untuk menguji validitas dan reliabilitas kuesioner, mengestimasi waktu pencacahan setiap responden, mengidentifikasi beban pencacah, uji coba pendaftaran rumah tangga (listing), serta menguji moda pengumpulan data dan desain survei yang digunakan. Survei Pendahuluan dilaksanakan di Kota Surabaya dan Kota Malang pada tanggal 4 - 7 Januari 2023. Kegiatan berikutnya adalah pencacahan lapangan yang dilaksanakan pada tanggal 23 Januari 2023 - 3 Februari 2023. Penelitian ini mengangkat topik “Kajian Pemanfaatan *Mobile Positioning Data* (MPD) dalam Menunjang Statistik Pariwisata” dengan mengambil lokasi penelitian di Kota Surabaya dan Kota Malang.

3.1.2 Cakupan Penelitian

Terdapat perbedaan cakupan yang digunakan untuk Survei Kajian MPD dan Survei Wisatawan Nusantara pada penelitian ini. Cakupan penelitian tersebut adalah sebagai berikut.



3.1.2.1 Survei Kajian MPD (SPRT.PKL62)

Berikut ini merupakan cakupan yang digunakan dalam Survei Kajian MPD (SPRT.PKL.62):

1. Populasi

Populasi adalah seluruh elemen yang dikumpulkan dengan karakter tertentu dan dapat dibedakan satu sama lain (Prasetyo dan Sukim, 2016). Populasi dari penelitian ini adalah seluruh pegawai BPS Kota Surabaya, pegawai BPS Provinsi Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya.

2. Populasi Target

Populasi target adalah indikator-indikator yang dicari dan terkait dengan tujuan penelitian (Prasetyo & Sukim, 2016). Populasi target dari penelitian ini adalah seluruh pegawai BPS Kota Surabaya, pegawai BPS Provinsi Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya yang pernah melakukan perjalanan wisata nusantara dalam satu tahun terakhir.

3. Populasi Survei

Populasi survei adalah populasi dalam survei yang meliputi populasi target (Prasetyo & Sukim, 2016). Populasi dari penelitian ini adalah seluruh pegawai BPS Kota Surabaya, pegawai BPS Provinsi Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI

Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya yang pernah melakukan perjalanan wisata nusantara dalam satu tahun terakhir.

4. Unit Observasi

Unit observasi adalah informasi dari unit yang diperoleh secara langsung maupun melalui responden tertentu (Prasetyo & Sukim, 2016). Unit observasi dari penelitian ini adalah pegawai BPS Kota Surabaya, pegawai BPS Provinsi Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya yang menggunakan provider Telkomsel dan/atau Indosat pada satu tahun terakhir yang bersedia digunakan datanya untuk penelitian.

5. Unit Sampling

Unit sampling adalah unit dasar pengambilan sampel berupa elemen maupun klaster (BPS, Pusdiklat). Unit sampling dari penelitian ini adalah pegawai BPS Kota Surabaya, pegawai BPS Provinsi Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya yang menggunakan provider Telkomsel dan/atau Indosat pada satu tahun terakhir yang bersedia digunakan datanya untuk penelitian.

6. Unit Analisis

Unit analisis adalah subjek penelitian atau sesuatu yang berkaitan dengan fokus/komponen yang diteliti (Arikunto, 2010). Unit analisis dari penelitian ini adalah pegawai BPS Kota Surabaya, pegawai BPS Provinsi



Jawa Timur, mahasiswa Tingkat III Tahun Akademik 2022/2023 Politeknik Statistika STIS, mahasiswa selain Tingkat III (Tingkat I, II, dan IV) Tahun Akademik 2022/2023 Politeknik Statistika STIS yang berdomisili DKI Jakarta dan Jawa Timur, serta dosen Politeknik Statistika STIS beserta anggota rumah tangganya yang pernah melakukan perjalanan wisata nusantara setahun terakhir yang menggunakan provider Telkomsel dan/atau Indosat pada satu tahun terakhir yang bersedia digunakan datanya untuk penelitian.

7. Kerangka Sampel

Kerangka sampel adalah seluruh unit yang dijadikan dasar penarikan sampel dalam populasi (BPS, Pusdiklat). Kerangka sampel dari penelitian ini adalah daftar pegawai BPS Kota Surabaya, Provinsi Jawa Timur pada tahun 2022, daftar pegawai BPS Provinsi Jawa Timur pada tahun 2022, daftar mahasiswa Politeknik Statistika STIS, dan daftar dosen Politeknik Statistika STIS.

3.1.2.2 Survei Wisatawan Nusantara (VRTW.62.C)

Berikut ini merupakan cakupan yang digunakan dalam Survei Wisatawan Nusantara (VRTW.62.C):

1. Populasi

Populasi dari penelitian ini adalah seluruh anggota rumah tangga di Kota Malang.

2. Populasi Target

Populasi target dari penelitian ini adalah seluruh anggota rumah tangga di Kota Malang yang minimal salah satu anggota rumah tangganya pernah melakukan perjalanan wisata nusantara selama Triwulan IV 2022.

3. Populasi Survei

Populasi dari penelitian ini adalah seluruh anggota rumah tangga di Kota Malang yang minimal salah satu anggota rumah tangganya pernah melakukan perjalanan wisata nusantara selama Triwulan IV 2022.

4. Unit Observasi

Unit observasi dari penelitian ini adalah anggota rumah tangga di Kota Malang yang pernah melakukan perjalanan wisata nusantara selama Triwulan IV 2022 dan telah divalidasi keberadaannya.

5. Unit Sampling

Unit sampling dari penelitian ini adalah rumah tangga di Kota Malang yang minimal salah satu anggota rumah tangganya pernah melakukan perjalanan wisata nusantara selama Triwulan IV 2022.

6. Unit Analisis

Unit analisis dari penelitian ini adalah anggota rumah tangga di Kota Malang yang pernah melakukan perjalanan wisata nusantara selama Triwulan IV 2022.

7. Kerangka Sampel

Kerangka sampel dari penelitian ini adalah daftar blok sensus dari Sensus Penduduk 2020 di Kota Malang, Provinsi Jawa Timur.

3.1.3 Alat Operasional

Dalam penelitian ini, digunakan beberapa alat operasional sebagai berikut.

3.1.3.1 Microsoft Excel

Spreadsheet adalah sebuah *array* baris dan kolom dengan pembaruan dan tampilan hasil yang otomatis. Hal ini juga telah diperluas dengan *library* fungsi matematika dan statistika, fasilitas grafik dan diagram serbaguna, *add-in* seperti Solver dari Microsoft Excel, antarmuka pengguna grafis yang menarik dan sangat fungsional, dan kemampuan untuk menulis kode dalam bahasa seperti Visual Basic for Applications milik Microsoft (Baker & Sugden, 2003).



Microsoft Excel dirancang untuk tujuan yang cukup teknis. Beberapa grafik yang tersedia di Excel antara lain adalah grafik kolom, grafik batang, grafik garis, grafik area, grafik permukaan, grafik lingkaran, grafik donat, grafik saham, grafik radar, dan diagram pencar. Fitur Excel Recommended Charts dapat menganalisis data yang dipilih dan menyajikan cara terbaik untuk memvisualisasikan data tersebut menurut pertimbangannya. Excel sering dianggap sebagai alat untuk menghitung angka dan dikaitkan dengan analisis data kuantitatif. Excel dapat menangani jumlah data yang besar, menyediakan banyak atribut dan memungkinkan berbagai teknik tampilan (Meyer & Avery, 2009).

3.1.3.2 Google Sheets

Google Sheets adalah salah satu aplikasi *spreadsheet* yang disediakan oleh Google sebagai bagian dari *suite* produktivitas mereka yang dikenal sebagai Google Workspace. Google Sheets dapat melakukan beberapa fitur statistik seperti *mean*, *median*, *mode*, *skewness*, *standard deviation*, *variance*, *z-scores*, *z-tests*, *t-tests* (*single*, *independent*, *paired*), *analysis of variance* (*one-way*, *repeated*, *two-way*), korelasi, regresi, dan *chi-square*. Fungsi-fungsi di Google Sheets menggunakan notasi standar yang dimulai dengan tanda "=" diikuti oleh fungsi, kemudian rentang data dalam tanda kurung (Kunicki et al., 2019). Nina Semczuk (2020) mengonfirmasi manfaat kolaborasi di Google Sheets serta menunjukkan seberapa sering Google Sheets diperbarui dan bekerja dengan baik dengan data *real-time*. Keunggulan lainnya adalah kemampuan riwayat akses Google Sheets yang dapat dengan mudah meninjau perubahan *spreadsheet* historis secara per menit dan melihat riwayat akun yang telah mengakses dan memodifikasi dokumen (Parra et al., 2021).

3.1.3.3 Google Colaboratory

Google Colaboratory atau sering disebut juga dengan Colab merupakan lingkungan pengembangan dari Jupyter Notebook dengan basis cloud yang memungkinkan untuk membangun, melatih, dan menguji model *machine learning* dan *deep learning*. Colab memiliki tiga keunggulan utama, yaitu tidak perlu konfigurasi, akses gratis pada GPU, dan mendukung kolaborasi dengan mudah.



Hanya dibutuhkan akun Google dan internet *browser* untuk menggunakan Colab sehingga Colab dirasa mudah digunakan (Allaam & Wibowo, 2021). Colab berfokus pada *tools* berbasis *cloud* dan gratis dengan tujuan penelitian. Colab mendukung hampir seluruh *library* yang dibutuhkan dalam pengembangan *artificial intelligence*. Pada dasarnya, Google Colaboratory sama seperti Jupyter Notebook tetapi dijalankan secara *online* dan sifatnya gratis (Brahma, 2020).

Notebook Colaboratory didasarkan pada Jupyter dan bekerja seperti objek Google Docs yaitu dapat dibagikan dan pengguna dapat berkolaborasi pada *notebook* yang sama. Colaboratory menyediakan *runtime* Python 2 dan 3 yang telah dikonfigurasi sebelumnya dengan pustaka *machine learning* dan *artificial intelligence* yang penting, seperti TensorFlow, Matplotlib, dan Keras. Mesin virtual di bawah *runtime* (VM) dinonaktifkan setelah jangka waktu tertentu, dan semua data serta konfigurasi pengguna hilang. Namun, notebook tetap dipertahankan dan dimungkinkan untuk mentransfer file dari hard disk VM ke akun Google Drive pengguna. Akhirnya, layanan Google ini menyediakan *runtime* yang dipercepat dengan GPU (Carneiro et al., 2018).

3.1.3.4 Python

Python merupakan bahasa pemrograman *open source* dan berlisensi gratis. Python termasuk bahasa pemrograman tingkat lanjut yang mudah gunakan dengan menggunakan struktur data yang efisien (Hasanuddin et al., 2020). Python didesain untuk meningkatkan kualitas perangkat lunak, *developer productivity*, *program portability*, dan *component integration* (Lutz, 2010). Python telah menjadi komunitas komputasi ilmiah dan analisis data yang besar dan aktif karena berbagai alasan sejarah dan budaya. Dalam beberapa tahun terakhir, dukungan Python untuk *library* (seperti Pandas dan Scikit-learn) yang semakin baik telah membuatnya menjadi pilihan populer untuk tugas analisis data. Jika digabungkan dengan kekuatan Python secara keseluruhan untuk rekayasa perangkat lunak tujuan umum, Python adalah pilihan yang sangat baik sebagai bahasa utama untuk membangun aplikasi data (McKinney, 2012). Library yang digunakan dalam



penelitian ini diantaranya adalah Pyspark, Datetime, Pandas, Numpy, Matplotlib, H3, Openpyxl, Plotly, OS, Scikit-learn.

3.1.3.5 Tableau

Tableau adalah salah satu *software* yang digunakan dalam aplikasi BI. Dalam pengaplikasian BI pada Tableau, akan dihasilkan visualisasi data yang interaktif dengan cara mengimpor dataset ke dalam Tableau dan membuat visualisasi data yang dapat mempermudah pemahaman suatu informasi. Tableau juga dapat mengumpulkan data dari berbagai sumber, seperti *spreadsheet*, *database*, *cloud data*, dan *big data* (Hartama, 2018). Tableau adalah alat visualisasi data yang dapat digunakan untuk membantu *library* menganalisis data. Tableau juga dapat menjadi alat penting bagi peneliti yang ingin melengkapi teks mereka dengan unsur visual yang menarik untuk membuat data dan analisis mereka mudah dipahami, tidak hanya oleh departemen mereka, tetapi juga oleh investor, khalayak akademik lainnya, dan masyarakat umum (Beard & Aghassibakes, 2021).

3.1.3.6 Kepler.gl

Kepler.gl adalah aplikasi berbasis web yang sangat efektif untuk eksplorasi visual data geolokasi dalam skala besar dan cocok untuk semua jenis data. Dibangun di atas deck.gl, Kepler.gl dapat menampilkan jutaan titik yang mewakili ribuan perjalanan dan melakukan agregasi spasial secara *real-time*. Dengan menampilkan data geospasial dalam antarmuka web tunggal, Kepler.gl membantu pengguna dengan cepat memvalidasi ide dan mendapatkan wawasan dari visualisasi tersebut. Dengan menggunakan Kepler.gl, pengguna dapat menggunakan CSV atau GeoJSON ke browser, memvisualisasikannya dengan lapisan peta yang berbeda, mengeksplorasi data dengan memfilter dan mengagregasikan, dan mengekspor visualisasi akhir sebagai peta statis atau video animasi (He, 2018). Alat ini adalah salah satu kerangka visualisasi milik Uber, dan dibuat untuk membuat peta berbasis data yang memungkinkan kita untuk mendapatkan wawasan dari lokasi dan memberikan hasil bisnis/ilmiah. Kepler.gl memungkinkan ilmuwan data untuk melihat beberapa aspek cuaca dalam satu



visualisasi. Ini memungkinkannya untuk mendapatkan hasil yang lebih baik dari analisis dan memberikan kesimpulan yang lebih baik kepada komunitas (Koch et al., 2018).

3.2 Metode Penelitian

3.2.1 Penjaminan Kualitas dan Pembangunan Algoritma *Family Grouping* dan *Double Counting*

Metode penelitian yang digunakan dalam penjaminan kualitas data dan pembangunan algoritma *family grouping* dan *double counting* adalah dengan pendekatan studi literatur. Studi literatur merupakan serangkaian kegiatan ilmiah yang dilakukan dengan mengumpulkan sejumlah informasi yang relevan dengan topik atau masalah yang akan atau sedang dikaji dengan kepustakaan sebagai sumber utama referensi (I Made & Cahyaningrum, 2020). Pada penelitian ini, dasar yang digunakan untuk membangun algoritma *family grouping* dan *double counting* berasal dari artikel, paper, jurnal penelitian, buku ataupun sumber lainnya. Pendekatan metode ini dapat memberikan pengetahuan terkait metode-metode yang bisa digunakan untuk melakukan *family grouping* dan mengatasi permasalahan *double counting* pada data MPD.

3.2.2 Penyajian Statistik Wisatawan Nusantara melalui Pendekatan MPD dan Survei Konvensional

Metode penelitian yang digunakan pada penyajian statistik wisatawan nusantara adalah metode penelitian kuantitatif yang memanfaatkan metode matematis, terutama dalam bentuk statistik, sebagai metode utama untuk mendeskripsikan hasil penelitian. Metode penelitian kuantitatif berlandaskan pada filsafat positivisme yang digunakan untuk meneliti populasi atau sampel tertentu dengan pengumpulan data melalui instrumen penelitian dan analisis data menggunakan statistik (Sugiyono, 2005). Pada penelitian ini, metode kuantitatif yang digunakan adalah metode survei. Metode survei digunakan untuk mendapatkan data yang terjadi pada masa lampau atau saat ini, tentang keyakinan, pendapat, karakteristik perilaku, hubungan variabel dan untuk menguji beberapa hipotesis tentang variabel sosiologis dan psikologis dari sampel yang diambil dari



populasi tertentu. Dengan menggunakan metode survei, penelitian ini diharapkan dapat memberikan gambaran terkait perjalanan wisata dan karakteristik wisatawan asal Kota Malang yang bepergian ke luar kota selama Triwulan IV 2022.

3.3 Metode Pengumpulan Data

Pengumpulan data dimulai dari proses *updating* dan wawancara secara *door to door*. Pada Survei Kajian MPD (SPRT.PKL62), dilakukan pencarian relawan MPD melalui proses wawancara yang dilakukan di Politeknik Statistika STIS, kantor BPS Kota Surabaya, dan kantor BPS Provinsi Jawa Timur sekaligus untuk mendapatkan data relawan MPD. Pada Survei Wisatawan Nusantara (VRTW.62.C), proses *updating* dan wawancara secara *door to door* dilakukan ke masing-masing rumah tangga sampel terpilih di Kota Malang. Survei Kajian MPD menggunakan instrumen kuesioner SPRT.PKL62 dengan rincian sebagai berikut:

1. Blok I tentang identitas wilayah
2. Blok II tentang ringkasan
3. Blok IIIA tentang keterangan petugas
4. Blok IIIB tentang keterangan pencacahan
5. Blok IV tentang catatan
6. Blok V tentang keterangan anggota rumah tangga
7. Blok VI tentang keterangan perjalanan anggota rumah tangga
8. Blok VII tentang keterangan perjalanan terakhir yang dilakukan anggota rumah tangga
9. Blok VIII tentang keterangan perjalanan terakhir yang dilakukan rumah tangga
10. Blok IX tentang keterangan ART pengguna Telkomsel dan Indosat

Survei Wisatawan Nusantara menggunakan instrumen kuesioner VRTW.62.C yang merupakan kuesioner integrasi dengan Riset 2. Berikut ini merupakan rincian blok-blok pertanyaan dalam kuesioner VRTW.62.C yang digunakan untuk kebutuhan analisis penelitian ini:

1. Blok I tentang keterangan tempat rumah tangga
2. Blok II tentang ringkasan
3. Blok III tentang keterangan petugas
4. Blok IV tentang catatan
5. Blok V tentang keterangan anggota rumah tangga
6. Blok VI tentang rincian banyak perjalanan anggota rumah tangga

Moda pengumpulan data yang digunakan adalah *Computer Assisted Personal Interviewing* (CAPI) atau wawancara dengan bantuan komputer/*smartphone*. Petugas lapangan menggunakan moda pengumpulan data dengan CAPI yang mampu menyimpan data meskipun adanya keterbatasan sinyal.

3.4 Metode Penarikan Sampel

3.4.1 Desain Sampling

3.4.1.1 Survei Kajian MPD (SPRT.PKL62)

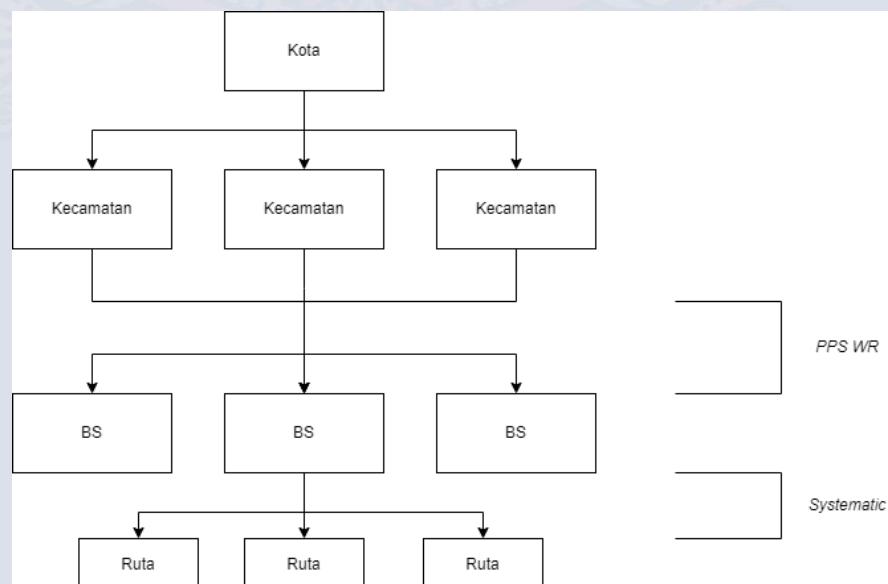
Metode penarikan sampel yang digunakan pada survei kajian MPD adalah *Judgment (Purposive) Sampling*. *Judgment (Purposive) Sampling* adalah metodologi pengambilan sampel secara acak dimana kelompok sampel yang ditargetkan memiliki atribut-atribut tertentu. Metode ini dapat digunakan pada banyak populasi, tetapi lebih efektif pada sampel yang berukuran kecil dan populasi yang homogen. Cara yang dilakukan pada *Judgment (Purposive) Sampling* adalah pengambilan sampel yang homogen, yakni memilih sekelompok orang dengan usia, jenis kelamin, latar belakang, atau pekerjaan yang sama. Pada penelitian ini, dipilih sampel dengan pekerjaan yang sama, yaitu orang yang bekerja di BPS Kota Surabaya dan BPS Provinsi Jawa Timur, dan sampel dengan latar belakang yang sama, yaitu orang yang sedang berkuliah di Politeknik Statistika STIS.

3.4.1.2 Survei Wisatawan Nusantara (VRTW.62.C)

Metode penarikan sampel yang digunakan dalam survei wisatawan nusantara adalah *Multistage Random Sampling* karena populasi yang heterogen dan berukuran sangat besar. *Sampling chart* yang digunakan dalam penelitian ini dapat terlihat pada gambar 3.1. Berikut tahapan penarikan sampel dengan Multistage Random Sampling:

1. Tahap Pertama: Melakukan pemilihan n blok sensus (BS) dari semua kecamatan secara *Probability Proportional to Size* (PPS WR-Systematic) dengan *size* jumlah keluarga di tiap blok sensus di Kota Malang hasil pemutakhiran Sensus Penduduk 2020.
2. Tahap Kedua: Melakukan pemilihan n rumah tangga dari blok sensus (BS) terpilih secara *Systematic Sampling* dengan variabel dasar pengurutan jumlah anggota rumah tangga yang minimal salah satu anggota rumah tangganya pernah melakukan perjalanan wisata selama Triwulan IV 2022.

Skema Sampling Survei Wisatawan Nusantara



Gambar 3.1 Skema Sampling

3.4.2 Teknik Estimasi Parameter Desain Sampling

3.4.2.1 Skema Sampling

Skema sampling yang diterapkan pada survei wisatawan nusantara yaitu sebagai berikut.

Tabel 3.1 Skema Sampling

Skema Sampling	Tahap	Unit	Banyak Unit		Metode Penarikan Sampel	Peluang Sampel	Fraksi Sampling
			dalam:				
		Populasi	Sampel				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prelist	1	Blok Sensus	N	n	PPS-WR	$\frac{X_i}{X}$	$n \frac{X_i}{X}$
	2	Rumah Tangga	M_i	m_i	Systematic Sampling	$\frac{1}{M_i}$	$\frac{m_i}{M_i}$

Keterangan:

h : Indeks blok sensus ($1, 2, \dots, n$)

N : Populasi blok sensus

n : jumlah sampel blok sensus

M_i : populasi rumah tangga blok sensus ke- i

m_i : jumlah sampel rumah tangga

X_i : jumlah rumah tangga blok sensus ke- i



3.4.2.2 Design Weight

1. Sampling Fraction

$$f_i = f_1 \times f_2 = \frac{nX_i}{X} \times \frac{m_i}{M_i} = \frac{n \times m_i \times X_i}{X \times M_i}$$

2. Initial Weight

$$\text{Weight} = w_i = \frac{1}{f_i} = \frac{X \times M_i}{n \times m_i \times X_i}$$

w_i merupakan penimbang unit rumah tangga perjalanan wisata dalam kondisi *response rate* 100%. Namun, kondisi di lapangan dapat terjadi *nonresponse* dalam pengumpulan data pelaku perjalanan wisata nusantara. Oleh sebab itu, diperlukan faktor koreksi untuk menghitung *initial weight* yang telah disesuaikan untuk *nonresponse*, yaitu:

$$w_i' = w_i \frac{1}{r_i} = \frac{X \times M_i}{n \times m_i \times X_i \times r_i}$$
$$r_{ij} = \frac{l_i}{l_i}$$

Keterangan:

w_i' : *design weight* dengan *non-response adjustment*

r_i : *response rate*

l_i : jumlah rumah tangga pelaku perjalanan wisata nusantara yang berhasil dicacah pada blok sensus ke-*i*

3.4.3 Penentuan Jumlah Sampel

Perhitungan ukuran sampel minimum pada populasi rumah tangga pelaku perjalanan wisata nusantara dengan metode Multiple Indicator Cluster Survey (MICS) versi 6 yang dikembangkan oleh United Nations Children's Fund (UN,2008) dengan formula sebagai berikut:

$$n = \frac{4r(1-r)deff}{(RME r)^2 pb AveSize RR}$$

Keterangan:

n : Ukuran minimum sampel yang dibutuhkan

r : Proporsi kejadian terhadap populasi (proporsi rumah tangga pelaku perjalanan wisata nusantara terhadap rumah tangga biasa)

RME : *Relative Margin of Error (10%)*

(Nilainya ditetapkan berdasarkan referensi dari SUTAS 2018)

$deff$: *Design Effect (=1,5)*

pb : Proporsi populasi target terhadap populasi penelitian

AveSize: Rata-rata jumlah anggota rumah tangga

RR : *Response rate*

3.5 Metode Analisis

3.5.1 Analisis Deskriptif

Analisis deskriptif adalah suatu analisis yang merupakan pengumpulan, pengolahan, dan penyajian serta interpretasi data secara kuantitatif atau persentase yang dapat disajikan dalam bentuk tabel atau grafik (Walpole, 1995). Menurut Sugiyono (2005), metode analisis deskriptif adalah suatu metode yang digunakan untuk menggambarkan atau menganalisis suatu hasil penelitian tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas. Dengan demikian, dari penjelasan di atas, metode analisis deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi. Metode analisis deskriptif digunakan untuk menganalisis hasil dari Survei Kajian MPD (SPRT.PKL62) dan Survei

Wisatawan Nusantara (VRTW.62.C). Adapun visual yang digunakan dalam analisis deskriptif pada penelitian ini antara lain:

3.5.1.1 Diagram Batang (*Bar Chart*)

Diagram batang merupakan diagram yang digunakan untuk melihat perbandingan jumlah perjalanan wisatawan nusantara atau jumlah wisatawan nusantara. *Centers for Diseases Control and Prevention* menyatakan bahwa diagram batang menampilkan variabel yang nilainya bersifat terbatas (*finite*) atau kontinu yang ditampilkan dalam pembagian berdasarkan variabel kategorik. Diagram ini digambarkan dengan sumbu horizontal dan vertikal dengan salah satu sumbunya berupa variabel kategorik. Setiap nilai dari kategori direpresentasikan dalam bentuk batang. Perbandingan antar kategori yang efisien dapat dilakukan dengan menggunakan visualisasi ini karena adanya perbedaan ketinggian/panjang dari setiap batang yang dapat memudahkan perolehan informasi secara pra-atentif. Untuk lebih mempermudah penyampaian informasi, pemberian warna pada batang tertentu dapat dilakukan dengan catatan tidak memberikan warna pada jumlah batang yang terlalu banyak (Mazza, 2009). Selain itu, diagram batang juga dapat berfungsi untuk melakukan pemeringkatan antara kategori dengan baik (Tableau Software, Inc., 2022).

3.5.1.2 Diagram Lingkaran (*Pie Chart*)

Pie chart atau grafik lingkaran adalah jenis grafik yang menunjukkan proporsi antara beberapa kategori dalam suatu dataset. Setiap kategori diwakili oleh sebuah *slice* pada lingkaran, di mana ukuran dari *slice* tersebut merepresentasikan persentase relatif dari data dalam kategori tersebut. *Pie chart* biasanya digunakan untuk menunjukkan bagian-bagian dari sebuah keseluruhan. Kelebihan dari *pie chart* adalah kemampuannya untuk dengan jelas menunjukkan bagian-bagian dari keseluruhan, sehingga sangat cocok untuk digunakan dalam situasi di mana perbandingan proporsi antara kategori sangat penting. Selain itu, *pie*

chart juga memiliki kemampuan untuk menarik perhatian pembaca dan mudah dipahami tanpa memerlukan pengetahuan khusus tentang statistik. Namun, salah satunya kelemahan *pie chart* adalah sulitnya untuk membandingkan ukuran dari beberapa *slice* yang sangat mirip ukurannya. Selain itu, jika terlalu banyak *slice*, maka *pie chart* dapat menjadi sulit untuk dibaca dan membungkungkan (Mazza, 2009).

3.5.1.3 Diagram Garis (*Line Chart*)

Diagram garis atau biasa disebut dengan *line chart* atau kurva merupakan visualisasi yang digunakan untuk menunjukkan nilai kuantitatif selama suatu periode waktu. Diagram ini terbentuk dari sistem koordinat yang terdiri dari sumbu vertikal dan horizontal. Sumbu-x menunjukkan rentang waktu data yang dipakai, sedangkan sumbu-y menunjukkan jumlah baris yang termasuk ke dalam indikator yang ingin dievaluasi. Setiap pasangan nilai dari kedua sumbu kemudian akan disambung dengan garis. Pada suatu publikasi yang berjudul *Visual Analysis Best Practices*, diagram garis baik digunakan untuk memantau tren dari waktu ke waktu (Tableau Software, Inc., 2022).

3.5.1.4 Dot Plot

Dot plot adalah salah satu jenis grafik yang digunakan untuk memvisualisasikan data numerik secara sederhana. Grafik ini terdiri dari serangkaian titik yang ditempatkan pada sumbu horizontal atau vertikal, di mana setiap titik mewakili satu observasi dalam data. *Dot plot* memiliki keunggulan dalam hal kemudahan interpretasi dan efektivitas visual dalam menunjukkan perbedaan antara kelompok data. *Dot plot* dapat digunakan untuk menampilkan perbedaan antara sampel, mengidentifikasi pola dalam data, serta memberikan pemahaman yang lebih baik tentang sebaran data. Selain itu, *dot plot* juga dapat memberikan informasi tambahan tentang data, seperti nilai minimum dan maksimum, serta nilai median dan kuartil.



Dot plot dapat diperluas dengan menggunakan warna dan ukuran titik untuk menunjukkan variabel kualitatif tambahan (Mazza, 2009).

3.5.2 Evaluasi Performansi

Evaluasi performansi adalah ukuran untuk mengevaluasi atau mengukur seberapa baik atau seberapa akurat model pengklasifikasian yang digunakan dalam memprediksi label kelas pada data yang ada (Han et al., 2012). Terdapat enam ukuran untuk mengukur model klasifikasi, yakni *accuracy*, *specificity*, *precision*, *recall*, nilai F1(*F1-score*), dan F- β (*F-Beta*). Akurasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi model klasifikasi secara keseluruhan (Nugroho et al., 2003). Semakin tinggi nilai dari akurasi berarti performansi model klasifikasi yang digunakan juga semakin baik.

Pada penelitian ini, pengukuran evaluasi kinerja model klasifikasi/algoritma menggunakan *confusion matrix* dan beberapa ukuran evaluasi yaitu *precision*, *recall*, *F1-score*, serta akurasi. Menurut Gorunescu (2011), *confusion matrix* merupakan alat ukur berupa matriks yang memberikan informasi rinci berupa jumlah ketepatan klasifikasi terhadap kelas dengan membandingkan hasil klasifikasi antara prediksi dengan kenyataan. Setiap kelas yang diprediksi memiliki empat kemungkinan *output* yang berbeda, yaitu:

- a. *true positive* (TP): hasil di mana model/algoritma dengan benar memprediksi kelas positif.
- b. *true negative* (TN): hasil di mana model/algoritma dengan benar memprediksi kelas negatif.
- c. *false positive* (FP): hasil di mana model/algoritma salah memprediksi kelas positif.
- d. *false negative* (FN): hasil di mana model/algoritma salah memprediksi kelas negatif.

Tabel 3.2 *Confusion Matrix*

		<i>Actual Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	TP	FP
	<i>Negative</i>	FN	TN

Confusion matrix tersebut digunakan untuk menghitung beberapa ukuran evaluasi yang dapat menilai kesalahan prediksi dalam mengukur kinerja model/algoritma klasifikasi yang telah dibangun, diantaranya:

1. *Precision*

Precision adalah rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precision* dapat dihitung dengan formula berikut:

$$\text{Precision} = \frac{\Sigma TP}{\Sigma TP + FP}$$

2. *Recall*

Recall adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* dapat dihitung dengan formula berikut:

$$\text{Recall} = \frac{\Sigma TP}{\Sigma TP + FN}$$

3. *F-measure (F1-Score)*

F-measure adalah ukuran kombinasi yang menggabungkan kedua ukuran *Precision-Recall*. *F-measure* yang biasa digunakan adalah *F1-Score*. *F1-Score* dapat dihitung dengan formula berikut:

$$F1score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{recall}}$$

4. Akurasi

Akurasi digunakan untuk mengukur seberapa tepat suatu *model//algoritma classifier* dalam mengklasifikasikan setiap data sesuai dengan kelasnya. Akurasi dapat dihitung dengan formula berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}$$

3.6 Definisi Operasional

Definisi operasional adalah definisi yang akan digunakan dalam penelitian riset ini. Definisi ini didapatkan dari hasil studi literatur yang kemudian dilakukan sedikit modifikasi untuk kepentingan tujuan riset.

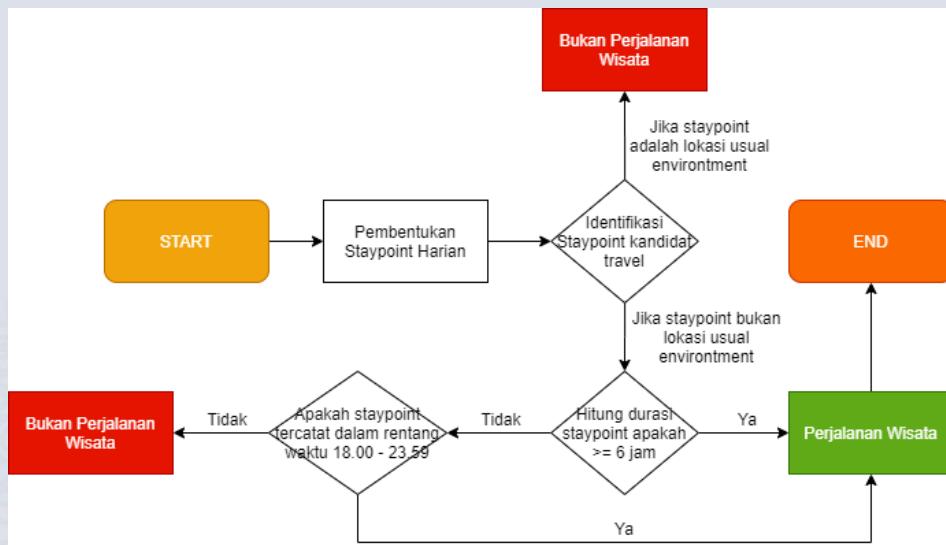
3.6.1 Survei Kajian MPD (SPRT.PKL62)

3.6.1.1 Perjalanan Wisata

Menurut IRTS (2008), perjalanan wisata adalah perjalanan yang dilakukan oleh seseorang yang keluar dari lingkungan kebiasaan selama kurang dari satu tahun dengan tujuan/maksud apapun (bisnis, rekreasi atau tujuan pribadi lainnya) selain untuk bekerja atau memperoleh penghasilan dari penduduk di tempat yang dikunjungi. Adapun menurut BPS (2021), Perjalanan Wisata (“Perjalanan”/”Bepergian”) adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 12 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/ gaji dari penduduk di tempat yang dituju) atau sekolah secara rutin. Tempat di luar lingkungan sehari-hari yang dimaksud meliputi berwisata ke objek wisata berpengelola, menginap di akomodasi komersial, dan berkunjung ke luar wilayah batas administratif dari *usual environment*. Menyesuaikan dengan konsep lingkungan kebiasaan (*usual environment*) dan juga algoritma penentuan trip melalui pendekatan MPD, maka terdapat ketentuan

tambahan, yakni penentuan lokasi lingkungan kebiasaan (*usual environment*) menggunakan batas geografis wilayah administratif setingkat kabupaten/kota (BPS, 2021), serta algoritma Penentuan Trip dengan MPD yang digunakan adalah sebagai berikut (BPS, 2021).

Flowchart Algoritma Penentuan Trip



Gambar 3.2 Flowchart Algoritma Penentuan Trip

Dengan beberapa penjelasan sebagai berikut:

- Staypoint* pada level kabupaten yang sudah menjadi kabupaten *usual environment* (*home, work, sirkuler bulanan, dan sirkuler mingguan*) tidak akan menjadi kandidat untuk *travel* (BPS, 2021).
- Staypoint* yang akan menjadi *travel* wisata adalah *staypoint* dengan durasi ≥ 6 jam atau *event staypoint* yang terjadi antara pukul 18.00 - 23.59 (mengasumsikan sedang menginap di tempat di luar *usual environment*).

Berdasarkan penjelasan tersebut, ditentukan konsep dan definisi perjalanan wisata yang digunakan dalam Survei Kajian MPD (SPRT.PKL62). Perjalanan Wisata adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat



di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 12 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/gaji dari penduduk di tempat yang dituju) atau sekolah secara rutin. Tempat di luar lingkungan sehari-hari adalah tempat di luar wilayah batas administratif dari lingkungan keseharian (*usual environment*) setingkat kabupaten/kota, yang meliputi berwisata ke objek wisata berpengelola, menginap di akomodasi komersial, dan berkunjung ke kabupaten/kota lain dengan tujuan/maksud apapun (bisnis, rekreasi atau tujuan pribadi lainnya).

3.6.1.2 Kegiatan Rutin

Konsep perjalanan wisata berlaku jika perjalanan yang dilakukan bukan termasuk kegiatan rutin. Kegiatan rutin adalah kegiatan sekolah dan atau bekerja (memperoleh upah/gaji sesuai tugas pokoknya dari penduduk di tempat yang dituju), yang dilakukan secara rutin, baik frekuensi, lokasi, maupun kegiatannya (BPS, 2021). Perjalanan apapun, tidak harus bekerja/sekolah, yang dilakukan secara rutin baik dari segi frekuensi, lokasi (walaupun terletak jauh dari *usual residence*), dan kegiatannya juga termasuk sebagai kegiatan rutin (IRTS, 2008), kecuali jika rutin berkunjung ke objek wisata dengan tujuan tidak bekerja maka termasuk sebagai perjalanan wisata.

3.6.1.3 Perjalanan/Trip

Sebuah perjalanan (trip) dapat terdiri dari beberapa kunjungan (*visits*) ke berbagai tempat, namun jumlah perjalannya tetap dihitung satu. Satu perjalanan dihitung mulai dari seseorang berangkat dari lingkungan kesehariannya ke suatu tempat tertentu hingga orang tersebut kembali ke lingkungan kesehariannya (IRTS, 2008).

3.6.1.4 Objek Wisata berpengelola

Objek wisata adalah segala sesuatu yang ada di daerah tujuan wisata yang merupakan daya tarik agar orang-orang mau datang berkunjung ke tempat tersebut, baik yang dikelola secara komersial maupun non komersial, seperti gunung, danau, sungai, pantai, laut, museum, benteng, situs peninggalan sejarah, dll (BPS, 2022). Objek wisata berpengelola adalah objek wisata yang memiliki pengelola usaha berupa sekelompok orang yang bertanggung jawab untuk mengendalikan dan mengelola objek wisata, atau sekelompok staff yang memiliki tugas tertentu untuk mengelola objek wisata. Pengelola objek wisata bisa merupakan pemerintah maupun masyarakat sekitar. Jika hanya membayar tiket parkir, maka tidak dihitung sebagai objek wisata berpengelola. Jika tidak terdapat tiket masuk di objek wisata tersebut, minimal terdapat pengelola. Berdasarkan Publikasi Statistik Objek Daya Tarik Wisata yang dikeluarkan oleh BPS (2019), berikut ini merupakan kategori tempat yang termasuk ke dalam objek wisata:



Kategori Objek Wisata dalam Survei Kajian MPD

KBLI *	<i>Deskripsi / Description</i>
(1)	(2)
91021	Museum yang dikelola pemerintah / <i>Government-run museum</i>
91022	Museum yang dikelola swasta / <i>Private-run museum</i>
91023	Peninggalan sejarah yang dikelola pemerintah / <i>Historical heritage managed by the government</i>
91024	Peninggalan sejarah yang dikelola swasta / <i>Historical heritage managed privately</i>
91025	Taman Budaya / <i>Cultural Park</i>
93221	Pemandian Alam / <i>Natural Baths</i>
93222	Wisata Gua / <i>Cave Tourism</i>
91031	Taman Konservasi Alam / <i>Nature Conservation Park</i>
91032	Taman Nasional (TN) / <i>National Parks</i>
91033	Taman Hutan Raya (TAHURA) / <i>Forest Park</i>
91034	Taman Wisata Alam (TWA) / <i>Nature Park</i>
91035	Suaka Marga Satwa (SM) dan Cagar Alam (CA) / <i>Wildlife and Nature Reserves</i>
91038	Hutan Lindung / <i>Protected forest</i>
91036	Taman Laut / <i>Sea Park</i>
91037	Taman Buru, Kebun Buru dan Areal Buru / <i>Buru Park, Buru Garden and Buru Area</i>
93223	Wisata Petualangan Alam / <i>Nature Adventure Tourist</i>
93231	Wisata Agro / <i>Agro tourism</i>
93233	Kolam Pemancingan / <i>Fishing Ponds</i>
93239	Daya Tarik Wisata Buatan/Binaan Manusia Lainnya / <i>Artificial Tourist Attraction</i>
68120	Kawasan Pariwisata / <i>Tourism area</i>
93210	Aktivitas Taman Bertema atau Taman Hiburan / <i>Amusement Parks</i>
93232	Taman Rekreasi / Taman Wisata / <i>Recreational Park</i>
93241	Arung Jeram / <i>Rafting</i>
93242	Wisata Selam / <i>Dive Tourism</i>
93243	Dermaga Marina / <i>Marina Dock</i>
93249	Wisata Tirta Lainnya / <i>Other Water Tourism</i>

10 -

Kajian Pemanfaatan Mobile Positioning Data (MPD) dalam Menunjang Statistik
Pariwisata

Gambar 3.3 Kategori Objek Wisata dalam Survei Kajian MPD

3.6.1.5 Akomodasi komersial

Akomodasi komersial adalah usaha akomodasi yang mewajibkan pengunjungnya/tamunya untuk membayar dengan besaran tertentu jika menggunakan/menginap di akomodasi tersebut, seperti hotel, hostel, villa, kos harian, dll.

3.6.1.6 Lama perjalanan wisata

Lama perjalanan wisata atau lama bepergian adalah waktu yang diperlukan wisatawan untuk pergi ke luar daerah sehari-harinya sampai kembali lagi, dihitung sejak meninggalkan rumah dan kembali ke rumahnya lagi. Lama perjalanan wisata dapat diukur dalam satuan hari, jam, atau menit. Dalam penelitian ini, lama perjalanan diukur menggunakan satuan jam dan dihitung dalam jam dengan pembulatan ke bawah atau durasi dalam jam yang terakhir.

3.6.1.7 Daerah Tujuan Utama

Daerah tujuan utama dari perjalanan wisata adalah tempat yang dikunjungi sebagai alasan melakukan perjalanan (IRTS, 2008). Namun, pada kondisi seseorang tidak dapat mengidentifikasi tempat tujuan utamanya, tempat tujuan utama ditentukan dengan kriteria sebagai berikut (IRTS, 2008):

1. Daerah tujuan utama adalah tempat yang paling lama dikunjungi.
2. Jika masih tidak dapat diidentifikasi, daerah tujuan utama adalah tempat yang dikunjungi dan paling jauh dari tempat tinggal keseharian.

3.6.1.8 Anggota Rumah Tangga (ART)

Dalam Survei Kajian MPD (SPRT.PKL62), untuk menyesuaikan dengan tujuan riset, definisi anggota rumah tangga dipersempit menjadi seseorang atau sekelompok orang yang terdaftar dalam Kartu Keluarga (KK) dan tinggal bersama dalam satu bangunan. Jika tidak memiliki KK, konsep keluarga mengacu pada UU No.52 tahun 2009. Berdasarkan UU 52 tahun 2009 tentang Perkembangan Kependudukan dan Pembangunan Keluarga, Keluarga adalah unit terkecil dalam masyarakat yang terdiri dari:



1. suami istri, atau
2. suami, istri dan anaknya, atau
3. ayah dan anaknya, atau
4. ibu dan anaknya.

Dengan demikian, anggota rumah tangga yang dimaksud adalah seseorang/sekelompok orang yang terdaftar dalam Kartu Keluarga (KK) dan tinggal serumah, atau seseorang/sekelompok orang yang menurut UU 52 tahun 2009 digolongkan menjadi keluarga dan tinggal serumah. Penentuan apakah seseorang ART sudah dapat dikatakan tinggal dalam rumah tangga tersebut adalah sebagai berikut:

1. Tercantum dalam KK, apabila tidak mempunyai KK, maka hanya anggota keluarga yang memiliki hubungan suami, istri, dan anak, serta telah menetap/tinggal di rumah tersebut selama 1 tahun secara berturut-turut. Hal ini karena dalam pengidentifikasi *usual environment* harus memperhitungkan dimensi geografis dan temporal selama 12 bulan ke belakang (BPS, 2021).
2. Jika anggota keluarga tersebut belum menetap/tinggal di rumah tersebut selama 1 tahun secara berturut-turut, penentuan tempat tinggal ditentukan berdasarkan dimana ia bertempat tinggal terlama dalam 1 tahun terakhir (1 Januari 2022 - 31 Desember 2022)

Definisi yang termasuk anggota keluarga yang digunakan dalam penelitian ini adalah anggota keluarga yang sudah tinggal 1 tahun atau lebih meskipun belum berniat untuk menetap (pindah datang). Sementara itu, definisi yang tidak termasuk anggota keluarga yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Anggota keluarga yang tinggal di tempat lain (luar BS) misalnya untuk sekolah atau bekerja, namun kembali ke orang tuanya seminggu sekali/lebih atau ketika libur telah dianggap memiliki lingkungan keseharian di tempat lain tersebut (luar BS).



2. Pembantu rumah tangga, tukang kebun atau sopir (yang tidak masuk ke dalam KK majikannya) meskipun tinggal dan makannya bergabung dengan rumah tangga majikan.
3. Orang yang menetap di pondok (yang tidak masuk ke dalam KK induk semangnya).
4. Kepala keluarga yang bekerja di tempat lain (tidak punya tempat tinggal tetap), tidak pulang setiap hari tapi pulang secara periodik (kurang dari 1 tahun) seperti pelaut, pilot, pedagang antar pulau, atau pekerja tambang.

3.6.1.9 Relawan

Relawan merupakan ART yang menggunakan kartu sim Telkomsel dan/atau Indosat dan menyetujui *agreement* atau perjanjian tertulis. Artinya, ART tersebut bersedia untuk dimintai nomor telepon dan bersedia data lokasinya selama 1 tahun terakhir digunakan untuk penelitian ini.

3.6.1.10 Handphone/Smartphone atau Telepon Seluler

Handphone/smartphone atau telepon seluler adalah perangkat telekomunikasi elektronik yang mempunyai kemampuan dasar yang sama dengan telepon tetap kabel, namun dapat dibawa ke mana-mana (*portable, mobile*), dan tidak perlu disambungkan dengan jaringan telekomunikasi kabel. Selain berfungsi sebagai telepon, telepon seluler modern biasanya mendukung layanan tambahan, seperti Short Messages Services (SMS), Multimedia Messages Service (MMS), *e-mail*, dan akses internet, aplikasi bisnis dan permainan, serta fotografi (BPS, 2022). Tablet, *smartwatch* (meskipun dapat digunakan untuk menelepon), dan telepon tetap nirkabel, tidak termasuk Telepon seluler/*Handphone/Smartphone* (BPS, 2022).



3.6.1.11 Unit *Handphone* Pribadi yang Digunakan

Handphone pribadi yang digunakan responden adalah unit *handphone* yang masih aktif dan memiliki paling sedikit 1 kartu sim aktif untuk penggunaan pribadi dimana responden secara relatif dapat menggunakan *handphone* tersebut secara penuh dan kontinu selama rentang waktu 1 Januari 2022 - 31 Desember 2022 dan apabila *handphone* yang dimaksud sempat tidak aktif dalam periode tersebut, maka periode waktu tidak aktifnya tidak lebih dari sebulan (berturut-turut). Perangkat *handphone* yang digunakan adalah yang benar-benar digunakan untuk keperluan komunikasi, tidak termasuk *handphone* yang hanya digunakan untuk jam, musik, atau *games*. Dengan kata lain jika fitur telepon tidak berfungsi/tidak bisa digunakan, maka tidak termasuk *handphone* (BPS, 2022).

3.6.1.12 Unit *Handphone* Utama yang Digunakan

Seperti yang sudah dijelaskan pada batasan penelitian, dengan tujuan untuk mengurangi kompleksitas pengolahan data dan penyederhanaan validasi individu yang termasuk *family grouping* dan *double counting*, penelitian ini memberlakukan batasan bahwa setiap individu diasumsikan hanya memiliki dan menggunakan satu *handphone* yang kemudian disebut dengan *handphone* utama. Untuk individu (ART) yang memiliki lebih dari satu *handphone*, perlu ditentukan salah satu *handphone* yang akan menjadi *handphone* utama dari ART tersebut. *Handphone* utama ditentukan berdasarkan urutan prioritas berikut:

1. Memiliki kartu sim aktif *provider* Telkomsel dan/atau Indosat (Karena penelitian ini menjalin kerjasama dengan BPS yang telah berkolaborasi dengan PT Telekomunikasi Selular (Telkomsel) dan PT Indosat Tbk (Indosat Ooredoo Hutchison) untuk memperoleh data MPD)

2. Paling sering digunakan atau paling banyak keterlibatannya dalam aktivitas sehari-hari (dibawa kemana-mana). (Dengan asumsi semakin sering *handphone* tersebut terlibat dalam aktivitas sehari-hari, maka semakin bervariasi aktivitas dan lokasi *record MPD* yang dihasilkan).
3. Waktu penggunaan terlama sejak pertama kali menggunakan *handphone* (*handphone* yang pertama kali dibeli/digunakan). (Dengan asumsi semakin lama seseorang menggunakan *handphone* tersebut, maka semakin banyak *record MPD* yang dihasilkan)
4. Jika tidak bisa ditentukan berdasarkan ketiga prioritas di atas, maka definisi *handphone* utama didasarkan pada pengakuan responden/ART.

3.6.1.13 Kartu sim Aktif

Kartu sim atau *sim card* adalah sebuah kartu pintar berukuran kecil yang dibenamkan pada *handphone* serta menyimpan informasi pengguna. Kartu sim merupakan salah satu produk yang dihasilkan oleh operator. Sesuai dengan batasan penelitian, bahwa periode analisis data MPD untuk melakukan *family grouping* dan mengatasi masalah *double counting* ditentukan selama setahun, yaitu sejak Januari hingga Desember 2022, maka ditetapkan definisi kartu sim aktif adalah kartu sim yang terpasang dalam *handphone* dan masih dapat digunakan untuk komunikasi pada periode waktu 1 Januari 2022 – 31 Desember 2022 dan apabila kartu sim yang dimaksud sempat tidak aktif dalam periode tersebut, maka periode waktu tidak aktifnya tidak lebih dari sebulan (berturut-turut). Kartu sim yang hanya berisi kuota internet juga termasuk dalam definisi kartu sim aktif yang dimaksud (BPS, 2022). *Sim card* yang telah melewati masa tenggang (sudah hangus), meskipun nomornya masih dapat digunakan sebagai nomor Whatsapp, dianggap tidak aktif, sehingga tidak disertakan dalam penghitungan jumlah *sim card* (BPS, 2022).



3.6.1.14 Provider/Operator

Provider atau operator adalah perusahaan/lembaga yang menyediakan jasa layanan telepon seluler (BPS, 2022), seperti Telkomsel, Indosat, XL Axiata, 3 (Three), dan Smartfren. Berikut ini merupakan nama-nama operator dan kartu sim yang termasuk ke dalamnya:

- a. Telkomsel : meliputi kartu Halo, Fleksi, As, Loop, Simpati, dan By.U
- b. Indosat : meliputi kartu StarOne, Matrix, mentari, dan IM3
- c. XI Axiata : meliputi kartu XL Prabayar, XL Prioritas, AXIS, dan Live.On
- d. 3 (Three) : meliputi kartu 3
- e. Smartfren : meliputi kartu Smartfren
- f. Lainnya : *provider* lainnya yang tidak disebutkan di atas seperti Esia dan sebagainya.

3.6.1.15 Jenis Kegiatan Utama

Kegiatan utama adalah kegiatan yang menggunakan waktu terbanyak dibandingkan dengan kegiatan lainnya dalam keseharian responden/ART pada periode waktu yang ditentukan (1 Januari 2022 - 31 Desember 2022). Dalam Statistik Wisatawan Nusantara 2019, BPS mengkategorikan kegiatan utama menjadi 4 kategori sebagai berikut:

1. Bekerja

Bekerja/berusaha adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus maupun kumulatif) dalam seminggu yang lalu (BPS, 2022). Sempinggu yang lalu pada penelitian ini merujuk kepada satu minggu terakhir dari periode analisis data MPD (24 Des – 31 Des 2022). Kegiatan tersebut termasuk pola kegiatan pekerja tak dibayar yang membantu dalam suatu usaha/kegiatan

ekonomi (BPS, 2022). Jika seseorang melakukan pekerjaan tetapi tidak bermaksud memperoleh atau membantu memperoleh penghasilan atau keuntungan maka dianggap tidak bekerja (BPS, 2022).

Menurut International Conference of Labour Statisticians (ICLS) (2013) dan SAKERNAS (2022), punya pekerjaan tetapi sementara tidak bekerja adalah keadaan dari seseorang yang mempunyai pekerjaan tetapi selama seminggu terakhir (24 Des – 31 Des 2022) sementara tidak bekerja karena berbagai sebab, seperti sakit, cuti, menunggu panenan, dan mogok bekerja termasuk sebagai bekerja/berusaha jika memenuhi beberapa syarat sebagai berikut:

- a. Masih tetap mendapatkan gaji atau upah.
- b. Masih memiliki keterikatan kerja yang ditetapkan dengan perkiraan kembali bekerja setelah waktu absen singkat tertentu.
- c. Absen dikarenakan alasan ketidakhadiran yang biasanya berdurasi pendek, dan di mana keterikatan kerja dipertahankan secara umum, meliputi:
 - Cuti sakit karena sakit atau cedera sendiri
 - Hari libur umum, liburan atau cuti tahunan
 - Masa cuti melahirkan atau paternitas menurut undang-undang
- d. Absen dengan alasan lain yang lebih panjang namun kurang dari tiga bulan seperti:
 - Cuti pendidikan
 - Pemogokan kerja
 - Merawat orang lain
 - Ketidakhadiran pribadi lainnya
 - Kecelakaan kerja



- Larangan bekerja
 - Pengurangan kegiatan ekonomi
 - Penangguhan pekerjaan (misalnya karena cuaca buruk, kerusakan mekanis, listrik atau komunikasi, kekurangan bahan baku atau bahan bakar)
- e. Apabila tidak ditemukan kepastian kembali bekerja dan absen lebih dari tiga bulan maka seseorang dianggap tidak bekerja.
- f. Pada kasus berusaha sendiri apabila tidak melakukan kegiatan ekonomi yang dimaksud (termasuk melakukan pekerjaannya tapi hanya untuk rumah tangganya dan bukan untuk mencari keuntungan) pada musim sepi maka tidak termasuk kategori bekerja.
- g. Untuk responden yang berusaha:
- Jika tidak bekerja ≤ 3 bulan namun masih memiliki keterikatan dengan usahanya itu, maka dianggap bekerja.
 - Jika sudah tidak bekerja lebih dari 3 bulan maka dianggap sudah berhenti bekerja, walaupun masih terikat dengan usahanya
- h. Untuk responden yang bekerja sebagai buruh:
- Jika tidak bekerja ≤ 3 bulan namun masih memiliki keterikatan dengan pekerjaan itu, maka dianggap bekerja.
 - Jika sudah tidak bekerja lebih dari 3 bulan maka dianggap sudah berhenti bekerja, namun jika masih menerima penghasilan dari pekerjaan tersebut maka dianggap bekerja.

- i. Responden pekerja bebas dan pekerja keluarga tidak boleh sementara tidak bekerja.

Jika responden/ART memiliki lebih dari 1 pekerjaan, maka pertanyaan seputar pekerjaan pada Survei Kajian MPD (SPRT.PKL62) merujuk kepada pekerjaan utama. Pekerjaan utama ditentukan berdasarkan urutan prioritas kriteria sebagai berikut (KBJI 2014):

- a. Jika seseorang hanya mempunyai satu pekerjaan, maka pekerjaan tersebut digolongkan sebagai pekerjaan utama.
- b. Bila pekerjaan yang dilakukan lebih dari satu, maka pekerjaan utama adalah pekerjaan yang dilakukannya dengan waktu terbanyak selama periode penelitian.
- c. Jika waktu yang digunakan sama, maka pekerjaan yang memberi penghasilan terbesar dianggap sebagai pekerjaan utama.

2. Sekolah/Kuliah

Sekolah/kuliah adalah kegiatan bersekolah di sekolah formal baik pada pendidikan dasar, pendidikan menengah atau pendidikan tinggi selama periode penelitian.

3. Mengurus rumah tangga

Mengurus rumah tangga adalah kegiatan seseorang yang mengurus rumah tangga tanpa mendapatkan upah, misalnya: ibu-ibu rumah tangga dan anaknya yang membantu mengurus rumah tangga. Sebaliknya pembantu rumah tangga yang mendapatkan upah walaupun pekerjaannya mengurus rumah tangga dianggap bekerja (BPS, 2022).

4. Kegiatan lainnya

Kegiatan selain bekerja, sekolah, dan mengurus rumah tangga. Termasuk didalamnya mereka yang tidak mampu melakukan kegiatan, seperti orang lanjut usia, cacat jasmani, dan



penerima pendapatan/pensiun yang tidak bekerja lagi (BPS, 2022).

Selain itu orang yang dikategorikan sebagai penganggur terbuka juga termasuk dalam kategori ini, diantaranya adalah:

- a. Mereka yang tak punya pekerjaan dan mencari pekerjaan.
- b. Mereka yang tak punya pekerjaan dan mempersiapkan usaha.
- c. Mereka yang tak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan.
- d. Mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja.

3.6.1.16 Moda Angkutan Utama

Moda angkutan adalah alat yang digunakan untuk berpindah tempat dari satu tempat ke tempat yang lain. Moda angkutan dirancang untuk mengangkut orang atau barang, tetapi sebagian besar Moda Angkutan dapat membawa kombinasi keduanya. Menurut BPS (2019), moda angkutan adalah sarana transportasi yang digunakan untuk melakukan perjalanan. Moda angkutan terdiri dari pesawat, kereta api, bus, angkutan air, *travel*, kendaraan pribadi, kendaraan sewa, dan lainnya. Moda angkutan *travel* yang dimaksud disini adalah bukan paket perjalanan yang diselenggarakan oleh Biro Perjalanan melainkan sejenis angkutan penumpang tertentu seperti Cipaganti, Xtrans, dsb. Sedangkan moda angkutan lainnya seperti sepeda motor, becak, dll. Berdasarkan IRTS (2008), moda angkutan utama ditentukan berdasarkan salah satu dari prioritas kriteria berikut, yaitu:

1. Jarak tempuh paling jauh (km atau mil)
2. Waktu berkendara paling lama
3. Biaya yang dikeluarkan paling besar



Dalam penelitian ini, moda angkutan utama dikategorikan secara berurutan sebagai berikut:

1. Pesawat
2. Kapal laut / angkutan air
3. Kereta api
4. Kendaraan roda 4 / bus
5. Sepeda motor
6. Lainnya

Apabila dari ketiga kriteria, moda angkutan utama masih belum dapat diidentifikasi, maka pilih moda angkutan utama dengan nomor urut paling kecil.

3.6.2 Survei Wisatawan Nusantara (VRTW.62.C)

3.6.2.1 Perjalanan Wisata

Menurut IRTS (2008), perjalanan wisata adalah perjalanan yang dilakukan oleh seseorang yang keluar dari lingkungan kebiasaan selama kurang dari satu tahun dengan tujuan/maksud apapun (bisnis, rekreasi atau tujuan pribadi lainnya) selain untuk bekerja atau memperoleh penghasilan dari penduduk di tempat yang dikunjungi. Adapun menurut BPS (2021), Perjalanan Wisata adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 12 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/ gaji dari penduduk di tempat yang dituju) atau sekolah secara rutin. Tempat di luar lingkungan sehari-hari dimaksud meliputi berwisata ke objek wisata berpengelola, menginap di akomodasi komersial, dan berkunjung ke luar wilayah batas administratif dari *usual environment*. Menyesuaikan dengan konsep *usual environment* dan algoritma penentuan trip melalui pendekatan MPD, maka terdapat ketentuan tambahan, yakni penentuan lokasi lingkungan kebiasaan (*usual*



(environment) menggunakan batas geografis wilayah administratif setingkat kabupaten/kota (BPS, 2021).

Berdasarkan penjelasan tersebut, maka ditentukan konsep dan definisi perjalanan wisata yang digunakan dalam Survei Wisatawan Nusantara (VRTW.C.62). Perjalanan Wisata adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 12 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/ gaji dari penduduk di tempat yang dituju) atau sekolah secara rutin. Tempat di luar lingkungan sehari-hari adalah tempat di luar wilayah batas administratif dari lingkungan keseharian (*usual environment*) setingkat kabupaten/kota, yang meliputi berwisata ke objek wisata berpengelola, menginap di akomodasi komersial, dan berkunjung ke kabupaten/kota lain dengan tujuan/maksud apapun (bisnis, rekreasi atau tujuan pribadi lainnya).

3.6.2.2 Kegiatan Rutin

Konsep perjalanan wisata berlaku jika perjalanan yang dilakukan bukan termasuk kegiatan rutin. Kegiatan rutin adalah kegiatan sekolah dan atau bekerja (memperoleh upah/gaji sesuai tugas pokoknya dari penduduk di tempat yang dituju) yang dilakukan secara rutin, baik frekuensi, lokasi, maupun kegiatannya (BPS, 2021). Perjalanan apapun, tidak harus bekerja atau sekolah, yang dilakukan secara rutin baik dari segi frekuensi, lokasi (walaupun terletak jauh dari *usual residence*), maupun kegiatannya juga termasuk sebagai kegiatan rutin (IRTS, 2008), kecuali jika rutin berkunjung ke objek wisata dengan tujuan tidak bekerja maka termasuk sebagai perjalanan wisata.

3.6.2.3 Perjalanan/Trip

Sebuah perjalanan (trip) dapat terdiri dari beberapa kunjungan (*visits*) ke berbagai tempat, namun jumlah perjalanannya tetap dihitung satu. Satu perjalanan dihitung mulai dari seseorang berangkat dari lingkungan kesehariannya ke suatu tempat tertentu hingga orang tersebut kembali ke lingkungan kesehariannya (IRTS, 2008).

3.6.2.4 Objek Wisata Berpengelola

Objek wisata adalah segala sesuatu yang ada di daerah tujuan wisata yang merupakan daya tarik agar orang-orang mau datang berkunjung ke tempat tersebut, baik yang dikelola secara komersial maupun nonkomersial, seperti gunung, danau, sungai, pantai, laut, museum, benteng, situs peninggalan sejarah, dll (BPS, 2022). Objek wisata berpengelola adalah objek wisata yang memiliki pengelola usaha berupa sekelompok orang yang bertanggung jawab untuk mengendalikan dan mengelola objek wisata, atau sekelompok staf yang memiliki tugas tertentu untuk mengelola objek wisata. Pengelola objek wisata bisa merupakan pemerintah maupun masyarakat sekitar. Jika hanya membayar tiket parkir, tidak dihitung sebagai objek wisata berpengelola. Jika tidak terdapat tiket masuk di objek wisata tersebut, minimal terdapat pengelola. Misalnya objek wisata, seperti pantai yang tidak dikelola oleh pihak manapun meskipun terdapat biaya retribusi parkir, tidak termasuk ke dalam objek wisata berpengelola. Kategori tempat yang termasuk ke dalam objek wisata berdasarkan Publikasi Statistik Objek Daya Tarik Wisata yang dikeluarkan oleh BPS (2019) terlihat pada gambar 3.5.

Kategori Objek Wisata dalam Survei Wisatawan Nusantara

KBLI *	Deskripsi / Description
(1)	(2)
91021	Museum yang dikelola pemerintah / <i>Government-run museum</i>
91022	Museum yang dikelola swasta / <i>Private-run museum</i>
91023	Peninggalan sejarah yang dikelola pemerintah / <i>Historical heritage managed by the government</i>
91024	Peninggalan sejarah yang dikelola swasta / <i>Historical heritage managed privately</i>
91025	Taman Budaya / <i>Cultural Park</i>
93221	Pemandian Alam / <i>Natural Baths</i>
93222	Wisata Gua / <i>Cave Tourism</i>
91031	Taman Konservasi Alam / <i>Nature Conservation Park</i>
91032	Taman Nasional (TN) / <i>National Parks</i>
91033	Taman Hutan Raya (TAHURA) / <i>Forest Park</i>
91034	Taman Wisata Alam (TWA) / <i>Nature Park</i>
91035	Suaka Marga Satwa (SM) dan Cagar Alam (CA) / <i>Wildlife and Nature Reserves</i>
91038	Hutan Lindung / <i>Protected forest</i>
91036	Taman Laut / <i>Sea Park</i>
91037	Taman Buru, Kebun Buru dan Areal Buru / <i>Buru Park, Buru Garden and Buru Area</i>
93223	Wisata Petualangan Alam / <i>Nature Adventure Tourist</i>
93231	Wisata Agro / <i>Agro tourism</i>
93233	Kolam Pemancingan / <i>Fishing Ponds</i>
93239	Daya Tarik Wisata Buatan/Binaan Manusia Lainnya / <i>Artificial Tourist Attraction</i>
68120	Kawasan Pariwisata / <i>Tourism area</i>
93210	Aktivitas Taman Bertema atau Taman Hiburan / <i>Amusement Parks</i>
93232	Taman Rekreasi / Taman Wisata / <i>Recreational Park</i>
93241	Arung Jeram / <i>Rafting</i>
93242	Wisata Selam / <i>Dive Tourism</i>
93243	Dermaga Marina / <i>Marina Dock</i>
93249	Wisata Tirta Lainnya / <i>Other Water Tourism</i>

Gambar 3.4 Kategori Objek Wisata dalam Survei Wisatawan Nusantara

3.6.2.5 Akomodasi Komersial

Akomodasi komersial adalah usaha akomodasi yang mewajibkan pengunjungnya/tamunya untuk membayar dengan besaran tertentu jika menggunakan/menginap di akomodasi tersebut, seperti hotel, hostel, villa, kos harian, dll.

3.6.2.6 Rumah Tangga Biasa dan Anggota Rumah Tangga (ART)

a. Rumah Tangga Biasa

Definisi rumah tangga biasa dan ART yang digunakan pada Survei Wisatawan Nusantara (VRTW.62.C) mengacu kepada definisi rumah tangga biasa dan ART pada SAKERNAS 2022. Rumah tangga biasa adalah seseorang atau sekelompok orang yang mendiami sebagian atau seluruh bangunan fisik atau sensus dan biasanya tinggal bersama serta makan bersama dari satu dapur. Yang dimaksud dengan satu dapur adalah pengurusan kebutuhan sehari-harinya dikelola menjadi satu. Satu rumah tangga dapat terdiri dari satu anggota rumah tangga atau lebih. Rumah tangga biasa umumnya terdiri dari bapak, ibu dan anak. Rumah tangga sering diartikan keluarga. Namun, satu keluarga tidak selalu satu rumah tangga atau sebaliknya. Kriteria rumah tangga lebih mempertimbangkan aspek satu pengelolaan urusan “dapur” atau kebutuhan sehari-hari. Rumah tangga yang dicatat dalam Survei Wisatawan Nusantara (VRTW.62.C) hanya rumah tangga biasa. Berikut merupakan yang termasuk rumah tangga biasa:

1. Seseorang yang menyewa kamar/sebagian bangunan sensus, tetapi makannya diurus sendiri;
2. Keluarga yang tinggal terpisah di dua bangunan sensus, tetapi makannya dari satu dapur, asal kedua bangunan sensus tersebut masih dalam Blok Sensus yang sama, maka dianggap sebagai satu rumah tangga;
3. Pemondok dengan makan (indekos) yang jumlah pemondoknya kurang dari 10 orang dengan makan, pemondok dianggap sebagai anggota rumah tangga induk semangnya. Jika yang mondok dengan makan 10 orang atau lebih, rumah tangga yang menerima pondokan dengan makan merupakan rumah tangga biasa, sedangkan yang



mondok dengan makan dianggap sebagai rumah tangga khusus;

4. Jika beberapa orang yang bersama-sama mendiami satu kamar dalam satu bangunan sensus atau bangunan fisik, walaupun pengelolaan makannya sendiri-sendiri, setiap kamar dianggap satu satunya rumah tangga biasa;
5. Pengurus asrama, pengurus panti asuhan, pengurus lembaga pemasyarakatan, dan sejenisnya yang tinggal sendiri maupun bersama anak istri, serta anggota rumah tangga lainnya dianggap rumah tangga biasa.

b. Anggota Rumah Tangga

Anggota Rumah Tangga adalah semua orang yang biasanya bertempat tinggal di rumah tangga terpilih, baik yang sedang berada di rumah pada waktu pendataan maupun yang sementara tidak berada di rumah. Berikut ini merupakan yang termasuk ART:

1. Bayi yang baru lahir.
2. Orang yang sudah tinggal 1 tahun atau lebih, meskipun belum berniat untuk menetap (pindah datang). Termasuk orang yang belum tinggal 1 tahun tetapi sudah meninggalkan rumahnya 1 tahun atau lebih.
3. Orang yang tinggal kurang dari 1 tahun tetapi berniat untuk menetap (pindah datang).
4. Pembantu rumah tangga, tukang kebun atau sopir yang tinggal dan makannya bergabung dengan rumah tangga majikan.
5. Kepala rumah tangga yang bekerja di tempat lain (tidak punya tempat tinggal tetap), tidak pulang setiap hari tapi pulang secara periodik (kurang dari 1 tahun) seperti pelaut, pilot, sopir antar pulau, pedagang antar pulau, pekerja tambang dsb.



6. Penduduk yang bersekolah SD/SMP/SMA atau yang sederajat, yang tinggal di tempat lain (kost, boarding school, pondok pesantren).

Berikut ini merupakan yang tidak termasuk ART:

1. Seseorang yang sudah bepergian 1 tahun atau lebih, meskipun belum jelas akan pindah.
2. Orang yang sudah pergi kurang dari 1 tahun tetapi berniat untuk pindah.
3. Pembantu rumah tangga yang tidak tinggal di rumah tangga majikan.
4. Penduduk yang kuliah (menempuh pendidikan di perguruan tinggi/diploma/ universitas) yang tinggal di tempat lain (misalnya: kost, asrama, kontrak, dll).
5. KRT/ART yang bekerja dan memiliki tempat tinggal tetap di tempat lain dan tidak pulang secara regular/rutin minimal sekali dalam seminggu.

Penentuan seseorang sebagai ART perlu memperhatikan beberapa catatan sebagai berikut:

1. Jika diketahui seorang suami mempunyai istri lebih dari satu, maka ia harus dicatat di salah satu rumah tangga istri yang lebih lama ditinggali. Bila diketahui lamanya tinggal bersama istri-istrinya sama maka ia dicatat di rumah istri yang paling lama dinikahi.
2. Orang yang bekerja (bukan KRT) di tempat lain (luar BS), tidak memiliki tempat tinggal tetap, dan tidak pulang setiap hari tapi pulang secara periodik (kurang dari 1 tahun) maka ia dicatat di rumah tangganya, seperti:



- a. Orang yang bekerja sebagai petani dan tinggal di saung, tetapi pulang ke tempat tinggalnya setiap bulan.
 - b. Orang yang bekerja sebagai buruh bangunan dan tinggal di bedeng/kontainer, tetapi pulang ke tempat tinggal sebulan sekali.
 - c. Orang yang bekerja sebagai sopir antarpulau dan tinggal di terminal atau di kendaraannya, tetapi pulang secara periodik (kurang dari 1 tahun).
3. Beberapa kasus untuk anggota rumah tangga yang bekerja, sekolah, dan kuliah:
 - a. Orang yang bekerja di luar BS dan pulang minimal seminggu sekali secara rutin, maka dicatat di tempat tinggal rumah tangganya (berlaku untuk Kepala Rumah Tangga dan Anggota Rumah Tangga). Sedangkan untuk yang pulang lebih dari seminggu dicatat di tempat tinggal saat bekerja.
 - b. Untuk anak-anak yang masih bersekolah SD, SMP, SMA (atau sederajat) yang bersekolah di tempat lain dan tidak tinggal bersama orang tuanya/rumah tangganya tetap dicatat di tempat tinggal orang tuanya/rumah tangganya. Termasuk yang bersekolah di *boarding school* atau pesantren. Jika anak yang masih bersekolah SD, SMP, SMA (atau sederajat) sejak lahir tidak tinggal dengan orang tuanya, misalkan sejak lahir sudah tinggal dengan neneknya, anak tersebut dicatat di rumah tangga neneknya.
 - c. Bagi yang sedang berkuliah (di universitas maupun sekolah dengan tingkat pendidikan di atas SMA) di luar BS dan tidak tinggal bersama dengan rumah

tangganya, maka dicatat di tempat tinggalnya saat berkuliahan.

3.6.2.7 Handphone/*Smartphone* atau Telepon Seluler

Handphone/smartphone atau telepon seluler adalah perangkat telekomunikasi elektronik yang mempunyai kemampuan dasar yang sama dengan telepon tetap kabel, namun dapat dibawa ke mana saja (*portable, mobile*), dan tidak perlu disambungkan dengan jaringan telekomunikasi kabel. Selain berfungsi sebagai telepon, telepon seluler modern biasanya mendukung layanan tambahan, seperti *Short Messages Services* (SMS), *Multimedia Messages Service* (MMS), *e-mail*, dan akses internet, aplikasi bisnis dan permainan, serta fotografi (BPS, 2022). Tablet, *smartwatch* (meskipun dapat digunakan untuk menelepon), dan telepon tetap nirkabel, tidak termasuk Telepon seluler/*Handphone/Smartphone* (BPS, 2022).

3.6.2.8 Unit *Handphone* Pribadi yang Digunakan

Handphone pribadi yang digunakan responden adalah unit *handphone* yang masih aktif dan memiliki paling sedikit 1 kartu sim aktif pada saat pencacahan untuk penggunaan pribadi dimana responden secara relatif dapat menggunakan *handphone* tersebut secara penuh dan kontinu (BPS, 2022). Perangkat *handphone* yang digunakan adalah yang benar-benar digunakan untuk keperluan komunikasi. Tidak termasuk *handphone* yang hanya digunakan untuk jam, musik, atau games. Dengan kata lain, jika fitur telepon tidak berfungsi/tidak bisa digunakan, tidak termasuk *handphone* (BPS, 2022).

3.6.2.9 Unit *Handphone* Utama yang Digunakan

Untuk individu (ART) yang memiliki lebih dari satu *handphone*, perlu ditentukan salah satu *handphone* yang akan menjadi *handphone* utama dari ART tersebut. *Handphone* utama ditentukan berdasarkan urutan prioritas berikut:



1. Paling sering digunakan atau paling banyak keterlibatannya dalam aktivitas sehari-hari (dibawa kemana-mana).
2. Waktu penggunaan terlama sejak kali pertama menggunakan *handphone* (*handphone* yang pertama kali dibeli/digunakan).
3. Jika tidak bisa ditentukan berdasarkan kedua prioritas di atas, definisi *handphone* utama didasarkan pada pengakuan responden/ART.

3.6.2.10 Kartu sim Aktif

Kartu sim atau *sim card* adalah sebuah kartu pintar berukuran kecil yang dibenamkan pada *handphone* serta menyimpan informasi pengguna. Kartu sim merupakan salah satu produk yang dihasilkan oleh operator. Kartu sim aktif yang dimaksud adalah *sim card* yang biasa digunakan dan masih aktif pada saat pencacahan. Aktif adalah terpasang dalam HP dan masih bisa digunakan untuk komunikasi meskipun pada saat pencacahan HP-nya dalam kondisi dimatikan. Termasuk juga *sim card* yang hanya berisi kuota internet (BPS, 2022). *Sim card* yang telah melewati masa tenggang (sudah hangus), meskipun nomornya masih dapat digunakan sebagai nomor Whatsapp, dianggap tidak aktif, sehingga tidak disertakan dalam penghitungan jumlah *sim card* (BPS, 2022).

3.6.2.11 Provider/Operator

Provider atau operator adalah perusahaan/lembaga yang menyediakan jasa layanan telepon seluler (BPS, 2022), seperti Telkomsel, Indosat, XL Axiata, 3 (Three), dan Smartfren. Berikut merupakan nama-nama operator dan kartu sim yang termasuk ke dalamnya:

- a. Telkomsel : meliputi kartu Halo, Fleksi, As, Loop, Simpati, dan By.U
- b. Indosat : meliputi kartu StarOne, Matrix, Mentari, dan IM3



- c. XL Axiata : meliputi kartu XL Prabayar, XL Prioritas, AXIS, dan Live.On
- d. 3 (*Three*) : meliputi kartu 3
- e. Smartfren : meliputi kartu Smartfren
- f. Lainnya : *provider* lainnya yang tidak disebutkan di atas seperti Esia dan sebagainya.



BAB IV

HASIL DAN PEMBAHASAN

Perkembangan teknologi informasi dan komunikasi telah berkembang pesat dalam beberapa dekade terakhir dan membawa pengaruh besar serta telah menjadi bagian penting dalam kehidupan sehari-hari manusia. Salah satu contoh teknologi tersebut adalah *handphone* atau telepon seluler. *Handphone* adalah alat telekomunikasi elektronik dua arah yang dapat dibawa kemanapun dan memiliki kemampuan untuk mengirimkan pesan suara. Selain berfungsi sebagai telepon, *handphone* juga biasanya mendukung layanan seperti SMS, akses internet, dan sistem pelacakan GPS. Menurut data Badan Pusat Statistik (BPS) dari hasil pendataan Survei Susenas 2021, sekitar 65,87 persen penduduk Indonesia sudah memiliki *handphone* pada tahun 2021. Persentase tersebut meningkat sebanyak 3,03 persen dari tahun sebelumnya. Tingginya penggunaan *handphone* ini mencerminkan iklim keterbukaan dan penerimaan masyarakat terhadap perkembangan teknologi.

Kelebihan *handphone* sebagai alat telekomunikasi *portable* yang dilengkapi dengan sistem GPS telah banyak dimanfaatkan sejak sekitar tahun 1990, salah satunya untuk melacak lokasi pengguna melalui *Mobile Positioning Data* (MPD). *Mobile Positioning Data* (MPD) adalah suatu dataset berkala besar mengenai catatan transaksi dan lokasi dari pelanggan operator seluler (*mobile network operator/MNO*) yang diproses dan disimpan di dalam suatu sistem (BPS, 2021). Sejak tahun 2016, BPS telah mulai mengeksplorasi pemanfaatan MPD sebagai salah satu sumber data dalam penyusunan statistik pariwisata secara umum.

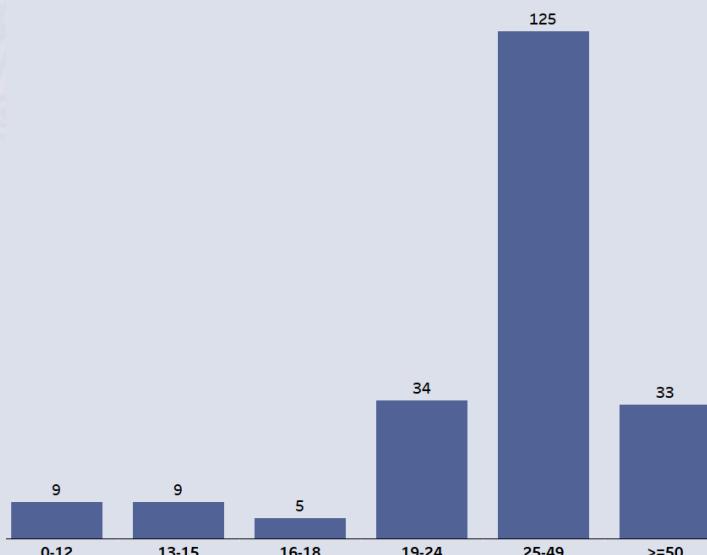
4.1 Gambaran Umum Karakteristik Relawan

Penelitian ini menggunakan data *Mobile Positioning data* (MPD) yang didapatkan dari *provider* Telkomsel dan Indosat. Data tersebut merupakan data yang didapatkan dari responden pengguna kartu Telkomsel dan/atau Indosat yang bersedia menjadi relawan untuk memberikan nomor telepon dan data MPD-nya melalui sebuah perjanjian tertulis (*agreement*). Berikut akan disajikan beberapa statistik mengenai relawan tersebut.

4.1.1 Jumlah Relawan berdasarkan Kepemilikan *Handphone* Pribadi

Sebelum dilakukan pengolahan data lebih lanjut, terlebih dahulu dilakukan eksplorasi jumlah relawan berdasarkan kepemilikan *handphone* pribadi. Eksplorasi ini dilakukan dengan mengingat bahwa data MPD sangat erat kaitannya dengan kepemilikan *handphone*. Jumlah relawan secara keseluruhan dapat terlihat pada gambar 4.1.

Jumlah Relawan Berdasarkan Kepemilikan *Handphone* Pribadi Menurut
Kelompok Umur Tahun 2022



Gambar 4.1 Jumlah Relawan Berdasarkan Kepemilikan *Handphone* Pribadi
Menurut Kelompok Umur

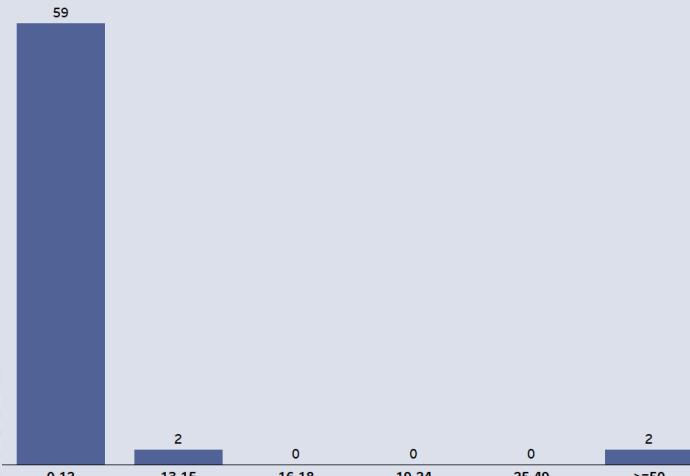


Berdasarkan gambar di atas, terlihat bahwa kategori usia 25—49 tahun memiliki jumlah relawan terbanyak, yakni 125 relawan, sedangkan kategori usia 16—18 tahun memiliki jumlah relawan yang paling sedikit, yaitu hanya sebanyak 5 relawan. Sebaran jumlah relawan atas kepemilikan ponsel pribadi didominasi oleh usia 19 tahun ke atas. Hal ini sesuai dengan hasil survei penggunaan TIK oleh Badan Litbang SDM Kominfo tahun 2017 yang mengatakan bahwa penggunaan *smartphone* paling banyak dari segi kelompok usia berada pada rentang usia 20—29 tahun, lalu diikuti oleh kelompok rentang usia 30—49 tahun. Selain itu pada rentang usia di bawah 19 tahun menunjukkan jumlah relawan yang paling sedikit atas kepemilikan *handphone* pribadi, hal ini sesuai dengan data Proporsi Individu yang Menguasai/Memiliki Telepon Genggam Menurut Kelompok Umur (Persen) yang dihitung oleh Badan Pusat Statistik pada tahun 2020 yang menyatakan bahwa persentase responden umur 15 tahun kebawah yang memiliki telepon genggam lebih sedikit dibandingkan dengan kelompok umur lainnya, yaitu hanya sebesar 24,96% saja.

4.1.2 Jumlah Anggota Keluarga yang Tidak Memiliki *Handphone* Pribadi

Sebelumnya dikatakan bahwa *handphone* telah menjadi bagian penting dalam kehidupan sehari-hari. Memiliki *handphone* sama pentingnya dengan memiliki pakaian, makanan, maupun tempat tinggal. Berkaitan dengan hal tersebut, menarik untuk diketahui apakah seluruh anggota keluarga relawan telah memiliki *handphone* pribadi. Dengan kata lain, akan dilihat apakah masih terdapat anggota keluarga relawan yang tidak memiliki *handphone* pribadi di era modernisasi ini.

**Jumlah Anggota Keluarga Relawan yang Tidak Memiliki *Handphone*
Pribadi Berdasarkan Kelompok Umur Tahun 2022**



Gambar 4.2 Jumlah Anggota Keluarga Relawan yang Tidak Memiliki *Handphone* Pribadi Menurut Umur

Dari hasil survei relawan MPD, terdapat 63 anggota keluarga relawan yang tidak memiliki *handphone*. Sebagian besar berasal dari kelompok umur 0—12 tahun, yaitu sebanyak 59 anggota keluarga. Fakta ini didukung oleh survei daring yang dilakukan oleh Statista Department Research pada tahun 2019 mengenai persepsi orang tua terhadap pemberian *smartphone* kepada anak-anak mereka di Indonesia. Dua jawaban terbanyak menyatakan bahwa sekitar 76% responden menyatakan bahwa mereka hanya memperbolehkan anak untuk menggunakan *smartphone* hanya jika berada di bawah pengawasan responden tersebut dan 45% responden khawatir jika anaknya mengalami kecanduan terhadap gawai. Pernyataan ini sejalan dengan sedikitnya jumlah relawan yang memiliki *handphone* pribadi di kelompok umur 0—12 tahun.

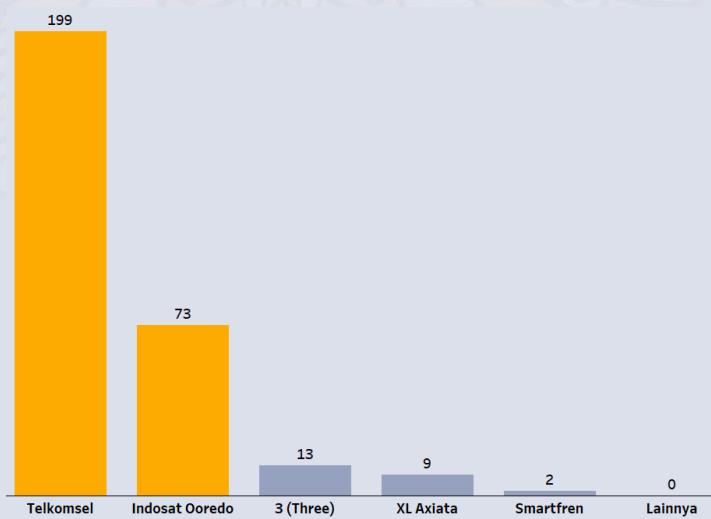
Pada survei MPD ini, dapat disimpulkan bahwa sebagian besar anggota keluarga relawan yang berumur 0—12 tahun atau yang masih mengenyam pendidikan sekolah dasar ke bawah belum memiliki *handphone*. Sementara itu, terdapat juga 2 anggota keluarga relawan yang tidak memiliki *handphone* pada masing-masing kelompok umur 13—15

tahun dan lebih dari 50 tahun. Sementara itu, anggota keluarga relawan pada kelompok umur lainnya tidak ada yang tidak memiliki *smartphone*.

4.1.3 Jumlah Kartu sim Aktif yang Digunakan Relawan

Salah satu bagian penting dari *handphone* adalah kartu sim. Kartu sim atau *sim card* adalah sebuah kartu pintar berukuran kecil yang dibenamkan pada *handphone* serta menyimpan informasi pengguna. Dengan adanya kartu sim aktif, pengguna dapat memanfaatkan layanan yang diberikan kartu sim tersebut, seperti telepon, SMS, internet, dan sebagainya. Selain itu, data MPD yang dimiliki operator/*provider* bisa dimanfaatkan dalam kajian pariwisata, salah satunya adalah untuk mengestimasi jumlah perjalanan wisata. Oleh karena itu, penting untuk mengetahui *provider* apa saja yang relawan gunakan. Penelitian ini telah melakukan analisa terhadap jumlah kartu sim aktif yang digunakan relawan berdasarkan jenis *provider*. Hasilnya dapat terlihat pada gambar 4.3.

Jumlah Kartu sim Aktif yang Digunakan Relawan pada *Handphone* Utama Berdasarkan *Provider* Tahun 2022



Gambar 4.3 Jumlah Kartu sim Aktif yang Digunakan Relawan pada *Handphone* Utama Berdasarkan *Provider*

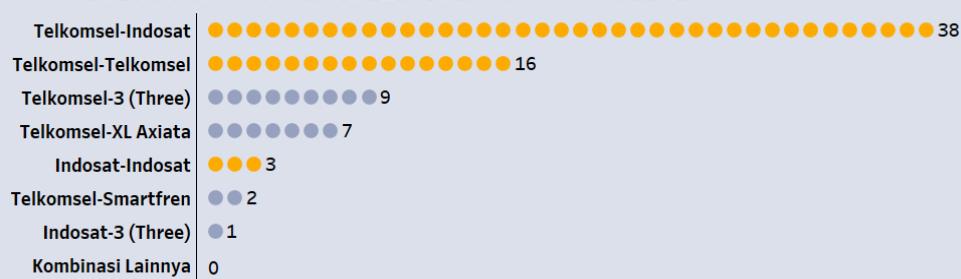
Grafik diatas menggambarkan jumlah kepemilikan kartu sim aktif yang digunakan relawan selama periode Januari hingga Desember tahun 2022 berdasarkan *provider*. Secara visual, dapat diidentifikasi bahwa sebagian besar relawan menggunakan kartu sim *provider* Telkomsel dan Indosat. Hal ini sesuai dengan batasan penelitian yang telah ditetapkan, yaitu berfokus kepada relawan yang menggunakan kartu sim Telkomsel dan Indosat. Berdasarkan grafik di atas, jumlah kartu sim Telkomsel adalah sebanyak 199 kartu, sedangkan jumlah kartu sim Indosat adalah sebanyak 73 kartu. Selain itu, tidak terdapat relawan yang menggunakan kartu sim aktif dengan *provider* lainnya (selain Telkomsel, Indosat, 3 (Three), XL Axiata, Smartfren). Selain *provider* yang menjadi fokus penelitian, sejumlah relawan juga menggunakan kartu sim dari *provider* lain, yakni 13 relawan menggunakan kartu 3 (three), 9 relawan menggunakan kartu XL Axiata, dan 2 relawan menggunakan kartu Smartfren.

4.1.4 Jumlah Relawan dengan Kepemilikan Ganda Kartu sim

Seiring dengan perkembangan teknologi, *handphone* berkembang dengan menyediakan dua slot kartu sim. Penggunaan dua kartu sim dalam satu *handphone* yang sama dapat menyebabkan terjadinya *double counting* dalam analisis data MPD. *Double counting* dapat diartikan sebagai individu yang sama dalam kepemilikan nomor telepon yang berbeda dalam satu *handphone*. Individu tersebut mungkin akan terdeteksi atau tercatat pada lokasi dan waktu yang berbeda untuk masing-masing nomor telepon. Hal tersebut disebabkan oleh jangkauan dan letak *tower* seluler (BTS) tiap *provider* serta aktivitas seluler untuk kedua nomor tersebut berbeda satu sama lain. Kondisi ini menyebabkan individu tersebut mempunyai *record* MPD yang berbeda antara nomor pertama dengan nomor kedua, sehingga seolah-olah dua *record* MPD yang berbeda tersebut merepresentasikan aktivitas seluler dari dua individu yang berbeda juga. Padahal dalam kenyataannya, dua *record* MPD tersebut

merujuk kepada individu yang sama. Oleh karena itu, untuk mendukung pemecahan masalah terkait *double counting* ini, dilakukan pendataan mengenai jumlah relawan yang menggunakan 2 kartu sim aktif pada satu *handphone* yang sama (*handphone* utama) agar dapat dilihat bagaimana penggunaan 2 kartu sim aktif tersebut dapat mempengaruhi hasil *Mobile Positioning Data* (MPD) dari relawan tersebut. Hasilnya dapat terlihat pada gambar 4.4.

Jumlah Relawan dengan Kepemilikan Ganda Kartu sim Aktif pada Handphone Utama Berdasarkan Kombinasi Provider Tahun 2022



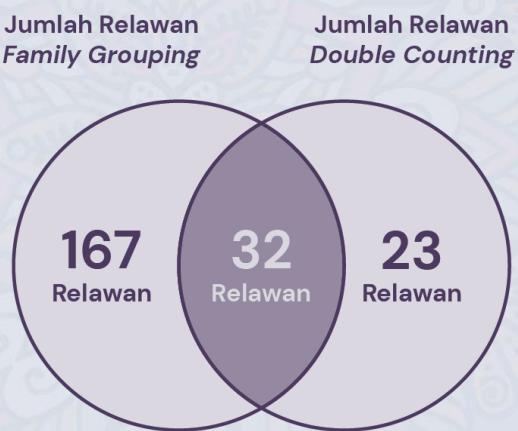
Gambar 4.4 Jumlah Relawan dengan Kepemilikan Ganda Kartu sim Aktif pada Handphone Utama Berdasarkan Kombinasi Provider

Grafik di atas memberikan informasi mengenai jumlah relawan yang menggunakan dua kartu sim aktif secara bersamaan dalam satu *handphone* utama. Kombinasi terbanyak dari kartu sim aktif ganda secara bersamaan adalah kartu sim Telkomsel-Indosat, yakni sebanyak 38 relawan, diikuti oleh Telkomsel-Telkomsel, Telkomsel-3 (Three), dan Telkomsel-XL Axiata. Sementara itu, kombinasi *provider* Indosat-3 (Three) memiliki jumlah yang paling rendah, yaitu hanya satu relawan. Pada survei ini, mayoritas relawan cenderung menggunakan kombinasi *provider* yang melibatkan Telkomsel sebagai salah satu penyedia layanan utama dalam penggunaan kartu sim ganda pada ponsel mereka.

4.1.5 Jumlah Relawan *Family Grouping* dan *Double Counting*

Dalam penelitian ini relawan dikelompokkan berdasarkan kriteria relawan yang sesuai untuk setiap algoritma yang ingin dibangun sesuai dengan tujuan penelitian. Relawan *double counting* adalah relawan yang menggunakan kombinasi kartu sim Telkomsel-Telkomsel, Indosat-Indosat, atau Telkomsel-Indosat dalam satu *handphone* yang sama. Sementara itu, relawan *family grouping* adalah relawan pengguna kartu sim Telkomsel dan/atau Indosat yang memiliki anggota keluarga yang juga merupakan relawan pengguna kartu sim Telkomsel dan/atau Indosat. Jumlah relawan beserta jumlah nomor pada masing-masing kelompok terlihat pada gambar 4.5 dan 4.6.

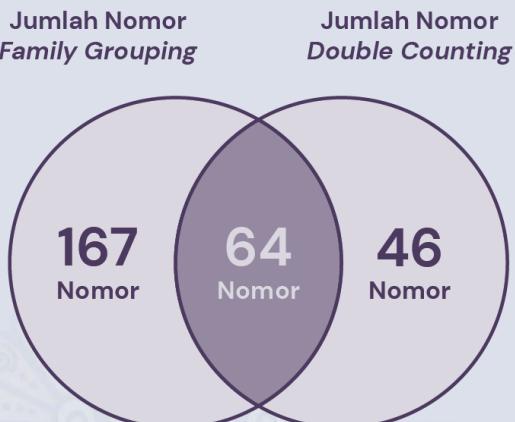
Jumlah Relawan *Family Grouping* dan *Double Counting*



Gambar 4.5 Jumlah Relawan *Family Grouping* dan *Double Counting*

Berdasarkan diagram di atas, terdapat sebanyak 167 relawan yang bersedia data MPD-nya digunakan dalam penelitian mengenai fenomena *family grouping* dan sebanyak 23 relawan yang bersedia data MPD-nya digunakan dalam penelitian mengenai permasalahan *double counting*. Selain itu, terdapat sebanyak 32 relawan yang bersedia data MPD-nya digunakan dalam penelitian mengenai keduanya, yakni fenomena *family grouping* dan permasalahan *double counting*.

Jumlah Nomor *Family Grouping* dan *Double Counting*



Gambar 4.6 Jumlah Nomor *Family Grouping* dan *Double Counting*

Dari jumlah relawan yang telah disebutkan sebelumnya, dihasilkan sebanyak 167 nomor yang digunakan untuk penelitian fenomena *family grouping* dan sebanyak 46 nomor yang digunakan untuk penelitian permasalahan *double counting*. Selain itu, terdapat sebanyak 64 nomor yang digunakan untuk meneliti keduanya, yaitu fenomena *family grouping* dan permasalahan *double counting*.

4.2 Penjaminan Kualitas Data (*Data Quality Assurance*)

Nomor *handphone* milik relawan selanjutnya diserahkan kepada PT Telekomunikasi Selular (Telkomsel) dan PT Indosat Tbk (Indosat Ooredoo Hutchison) melalui BPS untuk mendapatkan akses penuh data MPD yang digunakan dalam penelitian ini. Berikut adalah contoh data MPD yang diperoleh dari kedua perusahaan *provider* tersebut.

Tabel 4.1 Contoh Data MPD Telkomsel

msisdn	datetime	source	latitude	longitude	id_prov	id_kab	id_kec	event_month
msisdn00xxx	01/12/2021 00:57	chg	-6.35126	106.73384	36	74	30	2021-12
msisdn00xxx	01/12/2021 02:00	upcc	-6.35250	106.73833	36	74	30	2021-12

msisdn00xxx	01/12/2021 06:35	upcc	-6.35250	106.73833	36	74	30	2021-12
msisdn00xxx	01/12/2021 07:39	lba	-6.35250	106.73833	36	74	30	2021-12
msisdn00xxx	01/12/2021 12:36	upcc	-6.35250	106.73833	36	74	30	2021-12

Tabel 4.2 Contoh Data MPD Indosat

msisdn	datetime	source	latitude	longitude	prov	kab	kec	desa	month_event
msisdn00xxx	01/01/2021 00:02	LBA_ALL	-7.8049 43	110.357468	DI YOGYA-KARTA	YOGYA-KARTA	KRA-TON	KADI-PATEN	2021-01
msisdn00xxx	01/01/2021 00:03	LBA_ALL	-7.8049 43	110.357468	DI YOGYA-KARTA	YOGYA-KARTA	KRA-TON	KADI-PATEN	2021-01
msisdn00xxx	01/01/2021 00:03	LBA_ALL	- 7.80494 3	110.357468	DI YOGYA-KARTA	YOGYA-KARTA	KRA-TON	KADI-PATEN	2021-01
msisdn00xxx	01/01/2021 00:03	LBA_ALL	-7.8049 43	110.357468	DI YOGYA-KARTA	YOGYA-KARTA	KRA-TON	KADI-PATEN	2021-01
msisdn00xxx	01/01/2021 00:03	LBA_ALL	-7.8049 43	110.357468	DI YOGYA-KARTA	YOGYA-KARTA	KRA-TON	KADI-PATEN	2021-01



Dari tabel di atas, terlihat bahwa terdapat perbedaan antara data MPD *provider* Telkomsel dan data MPD *provider* Indosat. Perbedaan tersebut diantaranya adalah:

1. Tipe sumber data yang dihasilkan pada data MPD dari *provider* Telkomsel terdiri dari CHG, UPCC, dan LBA. Sementara itu, tipe sumber data dari *provider* Indosat hanya LBA. Terlihat bahwa tipe sumber data Telkomsel lebih beragam dibandingkan Indosat.
2. Banyak digit pada data koordinat *latitude* dan *longitude* pada data MPD *provider* Indosat lebih banyak dibandingkan Telkomsel, yakni 6 digit angka di belakang koma untuk data Indosat dan 5 digit angka di belakang koma untuk data Telkomsel.
3. Data MPD *provider* Telkomsel menggunakan id wilayah (seperti id provinsi, id kabupaten, dan id kecamatan) sebagai atribut identitas wilayah, sementara data MPD dari *provider* Indosat menggunakan nama daerah.
4. Cakupan identitas lokasi data MPD *provider* Indosat lebih lengkap, yakni mencakup provinsi hingga desa. Sementara itu, data MPD dari Telkomsel hanya mencakup provinsi hingga kecamatan.

Jejak digital yang ditinggalkan oleh pengguna pada data MPD sangat sensitif. Namun, jejak digital tersebut juga sangat penting karena menawarkan cara baru untuk mengukur dan melacak aktivitas spasial-temporal populasi. Di sisi lain, memanfaatkan pergerakan dan mobilitas individu untuk mendapatkan suatu informasi dapat membahayakan kebebasan bergerak dan hak pribadi. Oleh karena itu, masalah privasi harus ditangani dengan menggunakan mekanisme perlindungan privasi yang tepat, salah satunya adalah dengan membuat nomor telepon pelanggan menjadi anonim. Teknik tersebut biasanya disebut dengan teknik anonimisasi dan diterapkan pada tahap awal proses pemrosesan data (Putra et al., 2021). Oleh karena itu, data MPD yang diperoleh dari *provider* Telkomsel dan Indosat tidak lagi menampilkan

nomor asli relawan, melainkan sudah dalam bentuk *hashing* (anonim). Hal ini dilakukan sesuai dengan hukum dan peraturan telekomunikasi yang tidak mengizinkan untuk membagikan data pribadi kepada pihak lain tanpa persetujuan individu (Putra et al., 2021).

Pemrosesan data MPD melibatkan teknik *reverse geocoding* yang kemudian mampu menghasilkan alamat jalan, tempat deskriptif, atau wilayah administratif berdasarkan data lintang dan bujur (*latitude and longitude*). Kemampuan untuk mengkonversi menjadi kode geografis dari data koordinat MPD akan bermanfaat dalam analisis mobilitas masyarakat. Koordinat MPD yang diubah menjadi alamat atau wilayah administratif akan mempermudah proses untuk menemukan pergerakan dan interaksi yang disediakan oleh koordinat GPS (Putra et al., 2021).

Data MPD sendiri masih rentan terhadap berbagai masalah, seperti kesalahan manusia (*human errors*), baik yang bersifat sengaja maupun tidak disengaja; kesalahan transfer; *bug*, virus/*malware*, peretasan, dan ancaman siber lainnya; kegagalan perangkat keras atau *disk*; dan kompromi fisik terhadap perangkat atau telepon pelanggan (Putra et al., 2021). Oleh karena itu, diperlukan kontrol kualitas (*quality control*) dan pembersihan data (*data cleansing*). *Data cleansing* merupakan bagian yang penting dari proses pemrosesan data karena dapat meningkatkan kualitas data dan dapat meningkatkan produktivitas secara keseluruhan. Selain itu, tahap tersebut juga dapat mengeliminasi data yang salah atau sudah tidak relevan sehingga hanya menyisakan data berkualitas tinggi (Putra et al., 2021). Berikut adalah beberapa indikator *quality control* yang diterapkan terhadap data MPD Telkomsel dan Indosat.

4.2.1 Distribusi Jumlah Records

Distribusi Jumlah *Records Provider* Telkomsel



Gambar 4.7 Distribusi Jumlah *Records Provider* Telkomsel

Berdasarkan grafik di atas, terlihat bahwa distribusi jumlah *records* pada data Telkomsel setiap bulannya selama Desember 2021—November 2022 relatif datar (tidak ada yang terlalu tinggi dan tidak ada yang terlalu rendah). Hal ini menunjukkan bahwa kualitas data MPD Telkomsel cukup baik karena kemunculan *records* di setiap bulannya relatif stabil. Sementara itu, berdasarkan gambar 4.8, terlihat bahwa distribusi jumlah *records* pada data Indosat untuk setiap bulannya selama Desember 2021—November 2022 cukup berfluktuasi. Artinya, terdapat jumlah *records* yang sangat tinggi di bulan tertentu dan terdapat jumlah *records* yang sangat rendah di beberapa bulan lainnya. Hal ini menunjukkan bahwa kualitas data MPD Indosat kurang cukup baik karena kemunculan *records* di setiap bulannya berbeda-beda dan terdapat ketimpangan jumlah *records* antar bulan.

Distribusi Jumlah *Records Provider* Indosat



Gambar 4.8 Distribusi Jumlah *Records Provider* Indosat

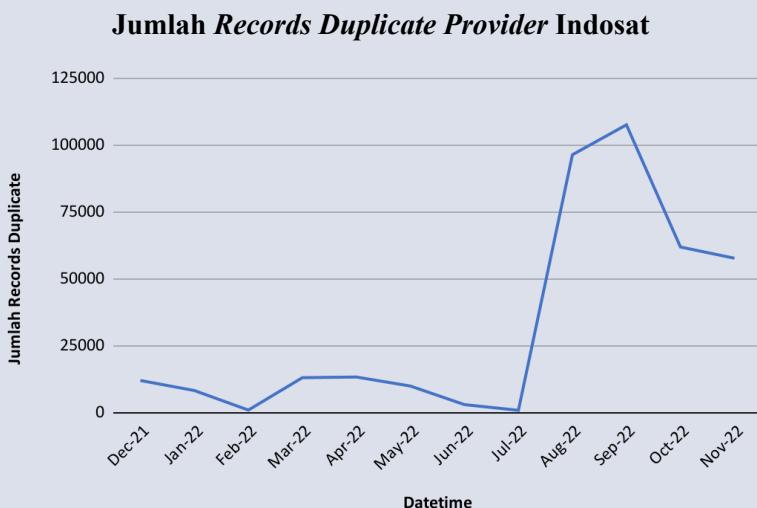
4.2.2 Jumlah *Records Duplicate*

Records duplicate yang dimaksud dalam penelitian ini adalah *records* yang memiliki kesamaan MSISDN, *datetime/timestamp*, *data source*, *latitude*, *longitude*, provinsi/id provinsi, kabupaten/id kabupaten, kecamatan/id kecamatan, dan *month event*.

Jumlah *Records Duplicate Provider* Telkomsel



Gambar 4.9 Jumlah *Records Duplicate Provider* Telkomsel



Gambar 4.10 Jumlah Records Duplicate Provider Indosat

Seperti yang terlihat pada gambar 4.9, selama periode Desember 2021—November 2022, jumlah *records duplicate* pada data *provider* Telkomsel menunjukkan angka nol. Hal tersebut memiliki arti bahwa *provider* Telkomsel tidak memiliki *record* yang duplikat sehingga dapat disimpulkan bahwa kualitas data MPD milik *provider* Telkomsel sangat baik. Sementara itu, dari gambar 4.10 dapat terlihat bahwa terdapat beberapa *records duplicate* pada data MPD *provider* Indosat dengan titik tertinggi pada bulan September. Artinya, jumlah *records duplicate* terbanyak ada pada bulan September 2022, yaitu lebih dari 100.000 *records duplicate*. Jumlah *records duplicate* terendah berada pada bulan Februari dan bulan Juli yang ditunjukkan dengan grafik yang mendekati titik nol. Sementara itu, pada bulan lainnya, jumlah *records duplicate* berada di atas angka nol. *Records duplicate* pada data MPD Indosat disebabkan oleh pembulatan koordinat latitude dan longitude menjadi 5 angka di belakang koma untuk menyesuaikan dengan tingkat ketelitian koordinat data MPD Telkomsel.

Untuk mengurangi beban komputasi, *records* duplikat pada data Indosat kemudian dihapus sehingga pada akhirnya hanya terdapat *records* unik saja pada data tersebut. sementara untuk *provider* Telkomsel, tidak

dilakukan hal yang sama karena pada data Telkomsel tidak terdapat *records duplicate*. Gambar 4.11 menunjukkan distribusi jumlah *records* setelah dilakukan eliminasi *records duplicate*.

Distribusi Jumlah Record Provider Indosat Setelah Record Duplicate Dihapus



Gambar 4.11 Distribusi Jumlah Record Indosat Setelah Record Duplicate Dihapus

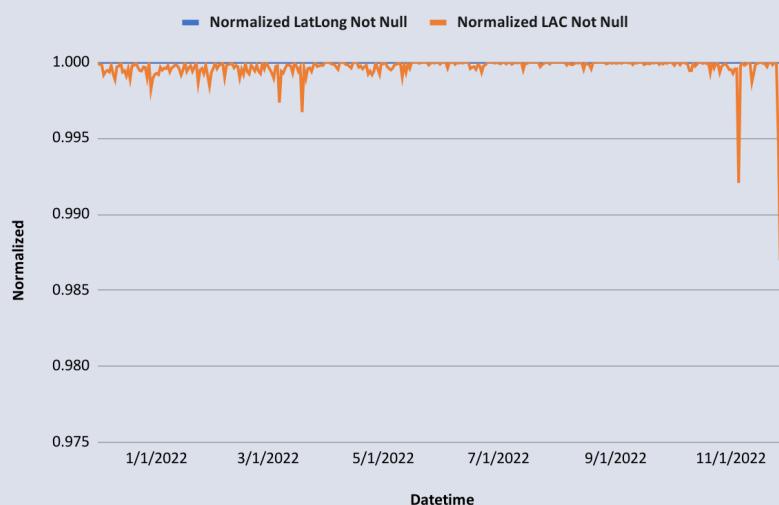
Dari gambar 4.8 dan 4.11, dapat terlihat perbedaan jumlah data sebelum dan sesudah dilakukan eliminasi *record duplicate*, terutama pada bulan Agustus dan September 2022. Setelah dilakukan penghapusan *records*, jumlah *records* pada bulan Agustus berkurang dari 916.472 menjadi 819.938 *records*. Sementara pada bulan September, jumlahnya berkurang dari 823.878 menjadi 716.181 *records*. Secara keseluruhan, jumlah records data provider Indosat berkurang dari 5.372.693 menjadi 4.985.544 *records*. Namun, distribusi jumlah *records* Telkomsel masih lebih stabil dibandingkan distribusi jumlah *records* Indosat meskipun telah dilakukan penghapusan *records duplicate*.

4.2.3 Latitude, Longitude, dan LAC Not NULL

Seperti yang sudah dijelaskan pada bagian sebelumnya, keterangan lokasi yang terdapat pada data MPD terdiri dari *latitude*, *longitude*, dan LAC. *Latitude* (garis lintang) adalah garis horizontal yang mengukur sudut antara suatu titik pada permukaan bumi dengan garis khatulistiwa.

Longitude (garis bujur) adalah garis vertikal yang mengukur sudut antara suatu titik yang dilewati garis bujur 0° atau 360° pada permukaan bumi. LAC (*location area code*) adalah kode numerik yang digunakan dalam jaringan seluler untuk mengidentifikasi suatu area di dalam suatu jaringan seluler. Data *latitude*, *longitude*, dan LAC yang bernilai NULL pada suatu *records* dapat menghambat proses pengidentifikasi lokasi *records* tersebut.

Normalized Latitude, Longitude, dan LAC Not Null Provider Telkomsel

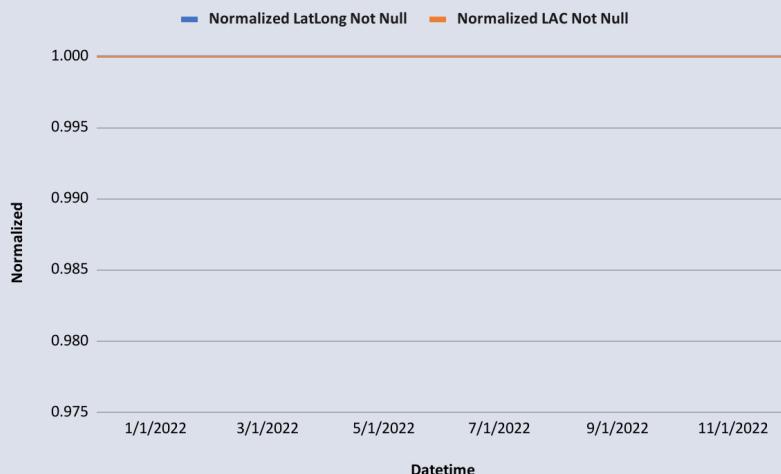


Gambar 4.12 Normalized Latitude, Longitude, dan LAC Not Null Provider Telkomsel

Berdasarkan data MPD *provider* Telkomsel selama periode Desember 2021—November 2022, terlihat bahwa hasil dari normalisasi atribut titik koordinat *latitude* dan *longitude* yang tidak mempunyai nilai *null* menunjukkan angka 1. Hal tersebut memiliki arti bahwa data koordinat *latitude* dan *longitude* dalam data *provider* Telkomsel terisi dengan lengkap (lengkap sempurna). Sementara itu, pada atribut LAC terlihat ada beberapa event_date yang nilai dari hasil normalisasi LAC-nya kurang dari 1, terutama pada event_date di bulan November 2022. Meskipun demikian, nilai tersebut masih lebih besar dari 0.9 yang menunjukkan bahwa data mengenai id provinsi, id kabupaten, dan id

kecamatan yang memiliki nilai *null* hanya sedikit atau dapat dikatakan LAC sangat lengkap. Oleh karena itu, dapat disimpulkan bahwa kualitas data MPD *provider* Telkomsel sangat baik. Sementara itu, hasil normalisasi untuk data *provider* Indosat dapat terlihat pada gambar 4.13.

Normalized Latitude, Longitude, dan LAC Not Null Provider Indosat



Gambar 4.13 Normalized Latitude, Longitude, dan LAC Not Null Provider Indosat

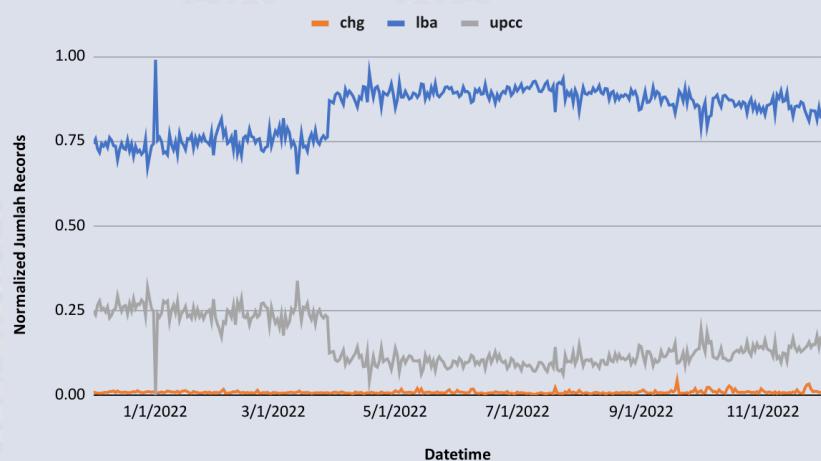
Dari gambar di atas, terlihat bahwa hasil dari normalisasi atribut titik koordinat *latitude* dan *longitude* beserta LAC *provider* Indosat selama periode Desember 2021 hingga November 2022 yang tidak memiliki nilai *null* menunjukkan angka 1. Hal tersebut menunjukkan bahwa data *latitude*, *longitude*, dan LAC pada *provider* Indosat terisi dengan lengkap secara keseluruhan atau dapat dikatakan lengkap sempurna. Hasil tersebut menunjukkan bahwa kualitas data MPD *provider* Indosat juga sangat baik.

4.2.4 Jumlah *Records* dan *Subscribers* per Hari Berdasarkan Tipe Data *Source*

Terdapat tiga data *sources* dari keseluruhan *records* dan *subscribers* yang ada pada data Telkomsel, yakni CHG, LBA, dan UPCC. CHG merupakan *billing* domain log yang berisi log transaksi, seperti panggilan, pesan, dan aktivitas lainnya. LBA merupakan data yang berisi

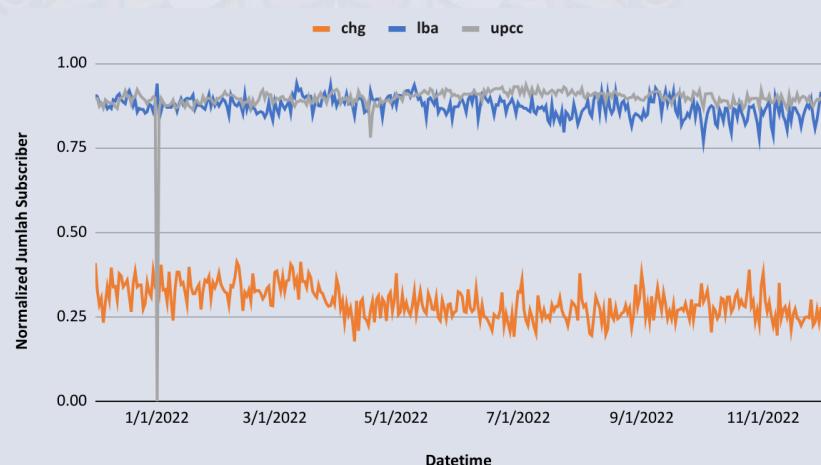
daftar lokasi *subscribers* dan menyediakan iklan khusus lokasi di perangkat seluler tersebut. UPCC merupakan catatan yang berisi log transaksi *subscribers* saat menggunakan internet. Jumlah *records* dan *subscribers* per hari berdasarkan tipe *data source* pada data Telkomsel terlihat pada Gambar 4.14 dan 4.15.

Jumlah Records per Hari Berdasarkan Tipe Data Source Provider Telkomsel



Gambar 4.14 Jumlah *Records* per Hari Berdasarkan Tipe *Data Source Provider* Telkomsel

Jumlah Subscribers per Hari Berdasarkan Tipe Data Source Provider Telkomsel



Gambar 4.15 Jumlah *Subscribers* per Hari Berdasarkan Tipe *Data Source Provider* Telkomsel

Berdasarkan gambar 4.14, terlihat bahwa *records* terbanyak adalah *records* dengan tipe *data source* LBA, sedangkan jumlah *records* yang paling sedikit adalah *records* dengan tipe *data source* CHG. Selain itu, terlihat juga bahwa terdapat perubahan jumlah *records* yang signifikan untuk *source* LBA dan UPCC, sedangkan CHG relatif konstan. Sementara itu, berdasarkan gambar 4.15, terlihat bahwa *subscribers* terbanyak pada data Telkomsel berasal dari *data source* LBA dan UPCC. Berbeda dengan jumlah *records*, terdapat perubahan jumlah *subscribers* yang signifikan untuk ketiga tipe *data source*.

Berbeda dengan data *provider* Telkomsel, data MPD *provider* Indosat hanya memiliki tipe *data source* LBA. Hal tersebut disebabkan oleh kebijakan pihak *provider* Indosat untuk menjaga kerahasiaan mengenai tipe *data source* yang dihasilkan dari aktivitas seluler penggunanya. Gambar 4.16 dan 4.17 menggambarkan jumlah *records* dan *subscribers* pada data *provider* Indosat. Kedua Grafik di bawah menunjukkan garis lurus pada angka 1 yang artinya terdapat jumlah *records* dan *subscribers* yang konstan untuk tipe *data source* LBA.

Jumlah Records per Hari Berdasarkan Tipe Data Source Provider Indosat



Gambar 4.16 Jumlah *Records* per Hari Berdasarkan Tipe *Data Source Provider* Indosat

Jumlah Subscribers per Hari Berdasarkan Tipe Data Source Provider

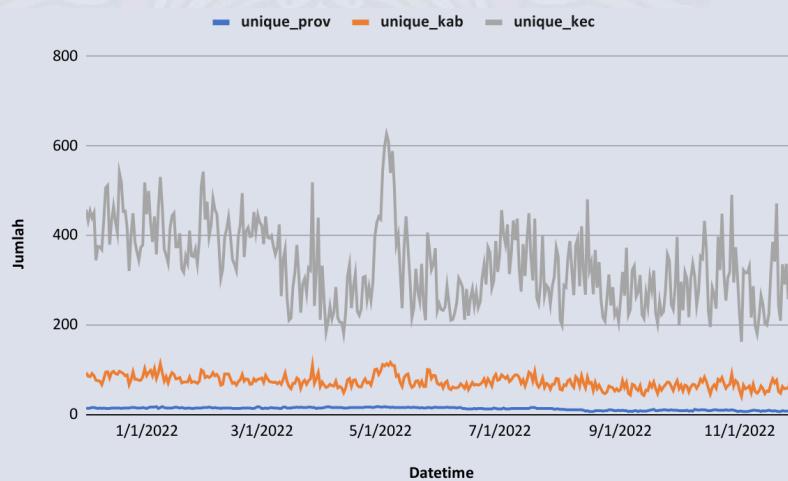
Indosat



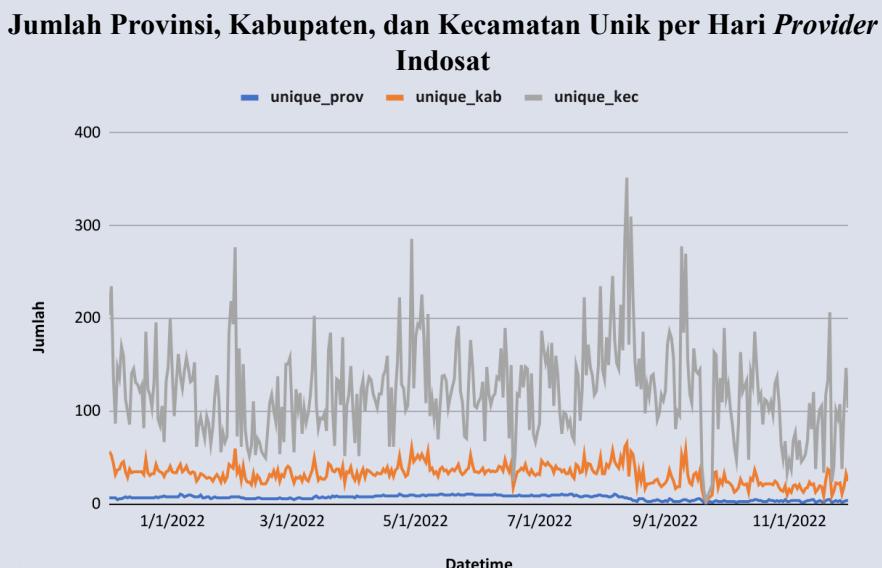
Gambar 4.17 Jumlah Subscribers per Hari Berdasarkan Tipe Data Source Provider Indosat

4.2.5 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari

Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari Provider Telkomsel



Gambar 4.18 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari Provider Telkomsel



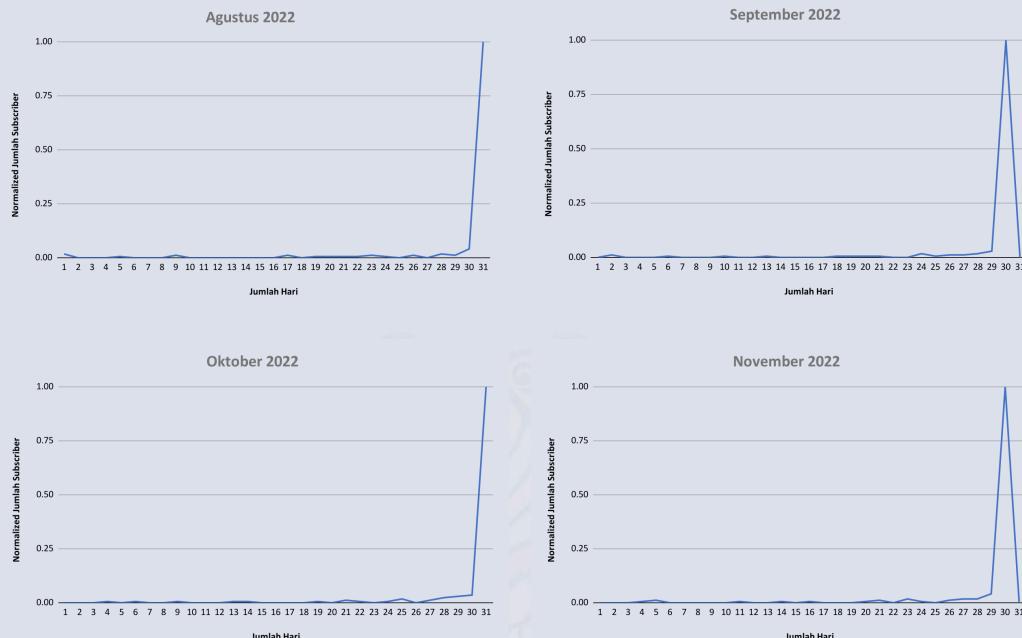
Gambar 4.19 Jumlah Provinsi, Kabupaten, dan Kecamatan Unik per Hari *Provider* Indosat

Seperti yang terlihat pada kedua gambar di atas, selama periode Desember 2021 hingga November 2022, jumlah provinsi, kabupaten, dan kecamatan unik pada data MPD Telkomsel dan Indosat umumnya berfluktuasi. Namun, fluktuasi yang terjadi pada data MPD Indosat terlihat lebih signifikan daripada data Telkomsel, khususnya pada data kecamatan unik. Secara keseluruhan, baik pada data MPD Telkomsel maupun Indosat, fluktuasi tertinggi terjadi pada jumlah kecamatan unik, kemudian disusul dengan jumlah kabupaten unik dan provinsi unik. Hal ini menunjukkan *subscriber* Telkomsel maupun Indosat pada tiap harinya melakukan mobilitas pada tingkat kecamatan (bepergian di dalam kota) dan pada tingkat kabupaten (bepergian ke luar kota) tapi masih dalam provinsi yang sama.

4.2.6 Jumlah Subscribers yang Hadir per Bulan

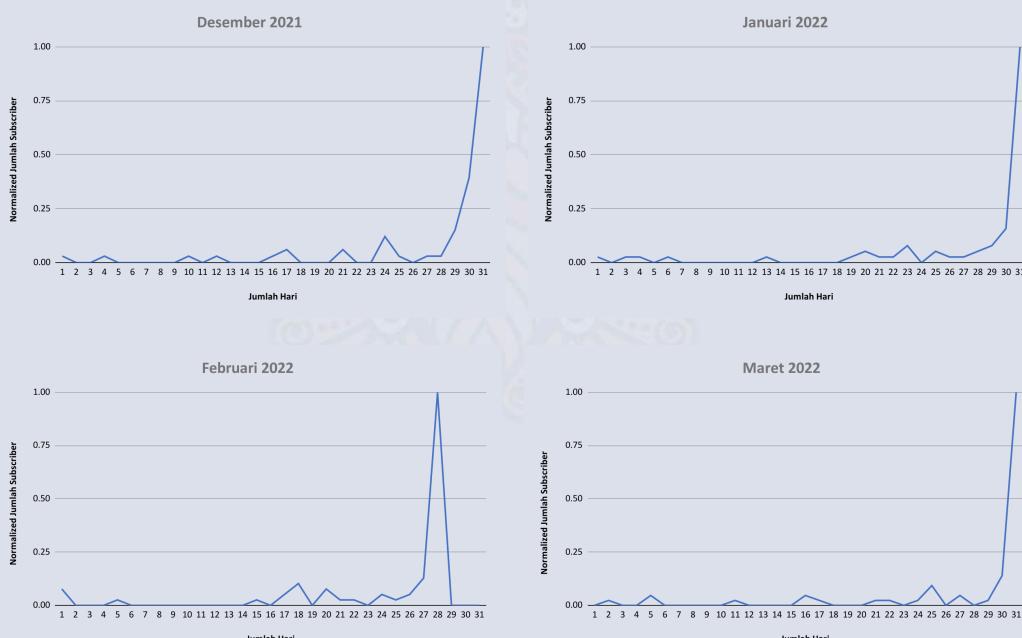
Jumlah Subscribers yang Hadir pada Data Provider Telkomsel





Gambar 4.20 Jumlah Subscribers yang Hadir Provider Telkomsel

Jumlah Subscribers yang Hadir pada Data Provider Indosat





Gambar 4.21 Jumlah Subscribers yang Hadir per Bulan *Provider* Indosat

Berdasarkan gambar 4.20 dan 4.21 terlihat bahwa pada periode Desember 2021 hingga November 2022, secara keseluruhan, normalisasi jumlah pengguna *provider* Telkomsel dan Indosat yang hadir masing-masing mencapai angka 1 pada hari-hari terakhir di setiap

bulannya. Hal tersebut membuktikan bahwa mayoritas pengguna *provider* hadir secara lengkap atau menyeluruh pada satu bulan tersebut. Akan tetapi, masih terdapat beberapa *spike* yang sangat kecil pada hari-hari tertentu yang menunjukkan bahwa masih terdapat beberapa pengguna yang tidak hadir secara lengkap atau menyeluruh di bulan tersebut.

4.2.7 Average Record per Hour Selama Satu Tahun

Average record per hour (rata-rata jumlah *records* per jam) menunjukkan rata-rata jumlah *records* tiap jam pada data MPD. *Average record per hour* dapat bervariasi tergantung pada berbagai faktor, diantaranya adalah aktivitas seluler pengguna, volume lalu lintas jaringan, dan waktu perekaman data. Atribut ini sangat berguna dalam memberikan wawasan mengenai penggunaan jaringan seluler dan perilaku pengguna.

Average Record per Hour Selama Satu Tahun Provider Telkomsel

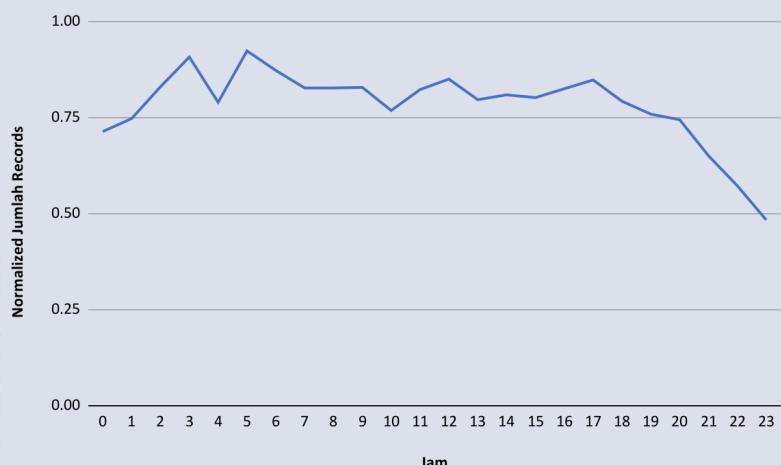


Gambar 4.22 Average record per hour Provider Telkomsel

Data MPD Telkomsel menunjukkan kenaikan rata-rata jumlah *records* secara signifikan pada pukul lima pagi. Rata-rata jumlah *records* berada pada titik tertinggi pada pukul 6 pagi hingga 20 (pukul 8 malam) yang artinya banyak transaksi yang terjadi pada jam tersebut. Banyaknya transaksi di jam tersebut dapat terjadi karena rentang waktu tersebut merupakan rentang waktu jam kerja sehingga terdapat banyak aktivitas

seluler dari pengguna Telkomsel. Sementara itu berdasarkan gambar 4.23, Sedikit berbeda dengan data *provider* Telkomsel, terlihat bahwa rata-rata jumlah *records* data MPD *provider* Indosat berada pada titik tertinggi pada pukul 3 pagi hingga 17 (pukul 5 sore).

Average Record per Hour Selama Satu Tahun Provider Indosat



Gambar 4.23 Rata-Rata Jumlah *Records* per Jam *Provider* Indosat

Secara keseluruhan, hasil dari penjaminan kualitas untuk data MPD provider Telkomsel dan Indosat adalah sebagai berikut.

Tabel 4.3 Hasil Penjaminan Kualitas Data MPD Provider Telkomsel dan Indosat

Penjaminan kualitas	Provider	Baik	Kurang Baik	Keterangan
Distribusi jumlah <i>records</i> (setelah <i>records duplicate</i> dihapus)	Telkomsel	✓		Distribusi cenderung berfluktuasi
	Indosat		✓	
Latitude, longitude, dan LAC <i>not null</i>	Telkomsel	✓		Hanya memiliki <i>record</i> bertipe data LBA
	Indosat	✓		
<i>Records</i> dan <i>subscribers</i> per hari berdasarkan tipe <i>data source</i>	Telkomsel	✓		Hanya memiliki <i>record</i> bertipe data LBA
	Indosat		✓	

Provinsi, kabupaten/kota, dan kecamatan unik	Telkomsel ✓
	Indosat ✓
Rata-rata jumlah <i>record</i> per jam	Telkomsel ✓
	Indosat ✓
Jumlah <i>subscribers</i> yang hadir per bulan	Telkomsel ✓
	Indosat ✓

Dapat disimpulkan bahwa kualitas data MPD dari kedua *provider* sudah baik. Namun, kualitas data MPD Telkomsel sedikit lebih baik dibandingkan dengan kualitas data MPD Indosat. Hal tersebut dapat dilihat berdasarkan distribusi jumlah *records* dan kelengkapan tipe *data source*.

4.3 Karakteristik *Data Training* dan *Data Testing*

Setelah dilakukan penjaminan kualitas, data MPD Telkomsel dan Indosat akan dibentuk menjadi pasangan nomor *family grouping* dan *double counting*. Pasangan nomor tersebut kemudian dibagi menjadi data *training* dan data *testing* dengan masing-masing proporsi 80% dan 20%. Data *training* digunakan untuk melatih algoritma *family grouping* dan *double counting* yang akan dibangun, *dataset* ini akan mempelajari pola hubungan antara fitur (*input*) dan target (*output*). Sementara itu, data *testing* digunakan untuk menguji kinerja algoritma *family grouping* dan *double counting* yang telah dibangun. Berikut adalah karakteristik dari data *training* dan *testing*.

Distribusi Jumlah Records Data Training Setelah Records Duplicate Dihapus

Gambar 4.24 Distribusi Jumlah *Records* Data *Training* Setelah *Records Duplicate* dihapus

Berdasarkan grafik di atas, terlihat bahwa distribusi jumlah *records* pada data *training* di tiap bulan dari Desember 2021—November 2022 relatif konstan (tidak ada yang terlalu tinggi dan tidak ada yang terlalu rendah). Hal ini menunjukkan bahwa kualitas data *training* tersebut sudah cukup baik karena kemunculan *records* di setiap bulannya relatif stabil.

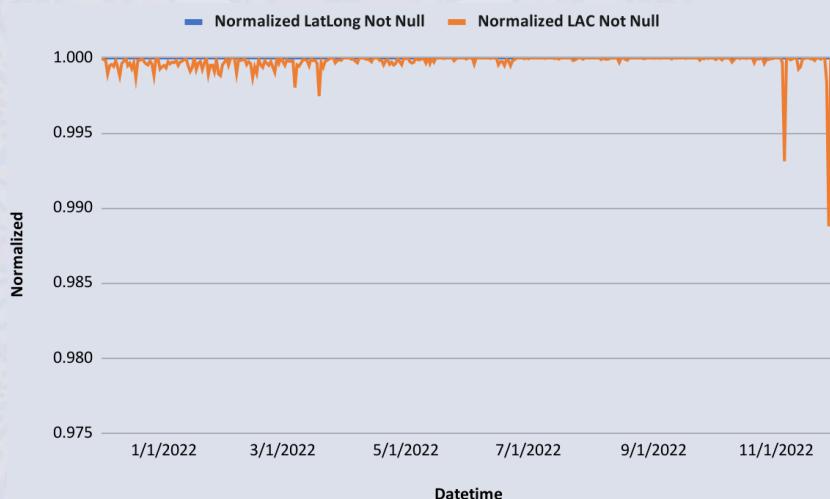
Distribusi Jumlah Records Data Testing Setelah Records Duplicate Dihapus

Gambar 4.25 Distribusi Jumlah *Records* Data *Testing* Setelah *Records Duplicate* dihapus

Sementara itu, berdasarkan grafik di atas, terlihat bahwa distribusi jumlah *records* pada data *testing* di tiap bulan dari Desember 2021—November 2022 berfluktuasi. Terdapat jumlah *records* yang sangat tinggi pada bulan Agustus sampai Oktober yang nilainya berbeda jauh dengan jumlah *records* di bulan lainnya, seperti pada bulan Juli dan November. Hal tersebut menunjukkan bahwa kualitas data *testing* tersebut tidak cukup baik karena kemunculan *records* di setiap bulannya berbeda-beda dan terjadi ketimpangan jumlah *records* antar bulannya.

Latitude, Longitude, dan Location Area Coordinate (LAC) Not NULL Data

Training Setelah Records Duplicate Dihapus

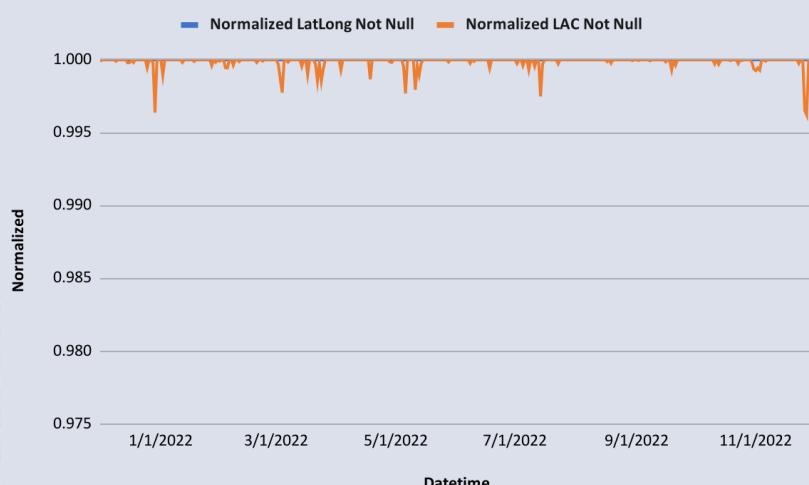


Gambar 4.26 Latitude, Longitude, dan LAC Not Null Data Training setelah Records Duplicate Dihapus

Berdasarkan gambar di atas, terlihat bahwa hasil dari normalisasi atribut titik koordinat *latitude* dan *longitude* yang tidak mempunyai nilai *null* menunjukkan angka 1. Hal tersebut memiliki arti bahwa data koordinat *latitude* dan *longitude* dalam data *training* tersebut terisi dengan lengkap. Sementara itu, pada atribut LAC terlihat ada beberapa hari yang nilai dari hasil normalisasi LAC-nya kurang dari 1, terutama pada hari-hari di bulan November 2022. Meskipun demikian, nilai tersebut masih lebih besar dari 0.9 yang menunjukkan bahwa data id provinsi, id kabupaten,

dan id kecamatan yang memiliki nilai *null* hanya sedikit. Oleh karena itu, dapat disimpulkan bahwa kualitas data *training* sangat baik.

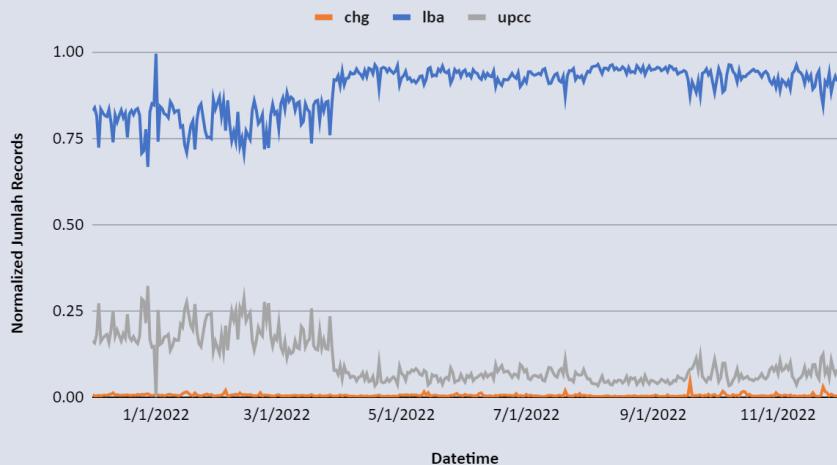
Latitude, Longitude, dan Location Area Coordinate (LAC) Not NULL Data
Testing Setelah Records Duplicate Dihapus



Gambar 4.27 Latitude, Longitude, dan LAC Not Null Data Testing Setelah Records Duplicate Dihapus

Hasil dari normalisasi atribut titik koordinat *latitude* dan *longitude* serta LAC untuk data *testing* selama periode Desember 2021 hingga November 2022 yang tidak memiliki nilai *null* menunjukkan angka 1. Hal tersebut menunjukkan bahwa data koordinat *latitude* dan *longitude* serta LAC untuk data *testing* terisi dengan lengkap secara keseluruhan. Sementara itu, pada atribut LAC terlihat ada beberapa hari yang nilai dari hasil normalisasi LAC-nya kurang dari 1. Meskipun demikian, nilai tersebut masih lebih besar dari 0,9 yang memiliki arti bahwa data mengenai id provinsi, id kabupaten, dan id kecamatan yang memiliki nilai *null* hanya sedikit. Oleh karena itu, dapat disimpulkan bahwa kualitas data *testing* sangat baik.

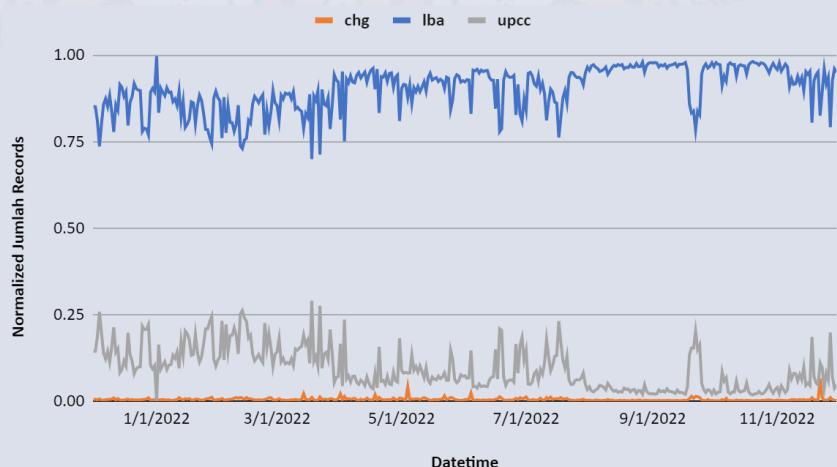
Jumlah Records per Hari Data Training Setelah Records Duplicate Dihapus



Gambar 4.28 Jumlah Records per Hari Data Training Setelah Records Duplicate Dihapus

Dari grafik di atas, terlihat bahwa jumlah *records* terbanyak adalah *records* bertipe *data source* LBA, sedangkan jumlah *records* paling sedikit adalah *records* bertipe *data source* CHG. Selain itu, terlihat juga bahwa terdapat perubahan jumlah yang signifikan untuk *records* bertipe *data source* LBA dan UPCC, sedangkan tipe *data source* CHG relatif konstan.

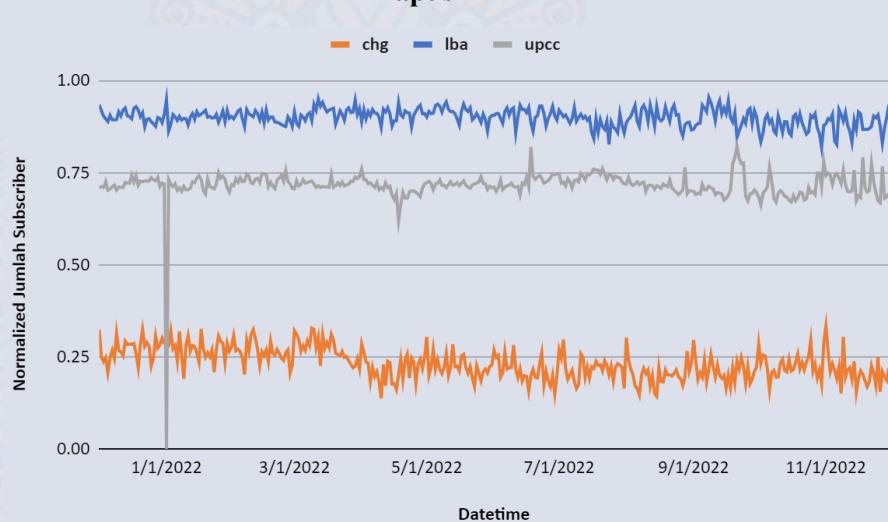
Jumlah Records per Hari Data Testing Setelah Records Duplicate Dihapus



Gambar 4.29 Jumlah Records per Hari Data Testing Setelah Records Duplicate Dihapus

Tidak jauh berbeda dari data *training*, berdasarkan grafik di atas, terlihat bahwa jumlah *records* terbanyak pada data *testing* adalah *records* bertipe data *source* LBA, sedangkan jumlah *records* yang paling sedikit adalah *records* bertipe data *source* CHG. Namun, terlihat juga bahwa jumlah *records* untuk tipe *data source* CHG relatif lebih konstan dibandingkan dengan *records* bertipekan *data source* LBA dan UPCC.

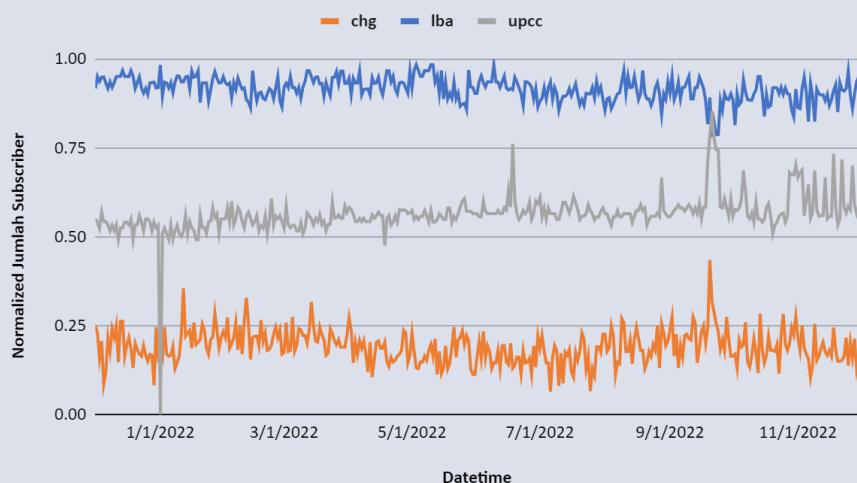
Jumlah Subscribers per Hari Data Training Setelah Records Duplicate Dihapus



Gambar 4.30 Jumlah Subscribers per Hari Data Training Setelah Records Duplicate Dihapus

Dari grafik di atas terlihat bahwa jumlah *subscribers* terbanyak pada data *training* adalah *subscribers* dengan *records* bertipe *data source* LBA sedangkan jumlah *subscribers* paling sedikit adalah *subscribers* dengan *records* bertipe *data source* CHG. Selain itu, terlihat juga adanya perubahan jumlah yang signifikan untuk *subscribers* dengan *records* bertipe *data source* UPCC, sedangkan LBA dan CHG relatif konstan.

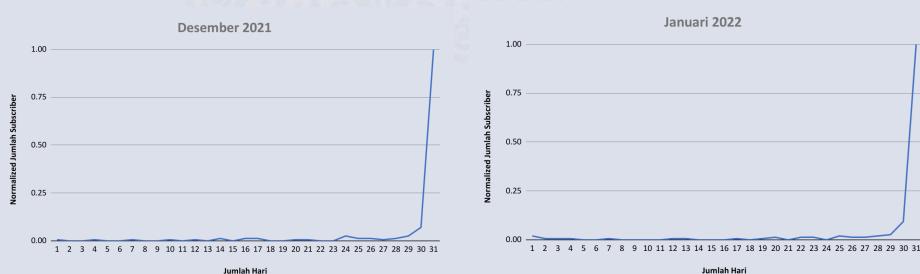
Jumlah Subscribers per Hari Data Testing Setelah Records Duplicate Dihapus

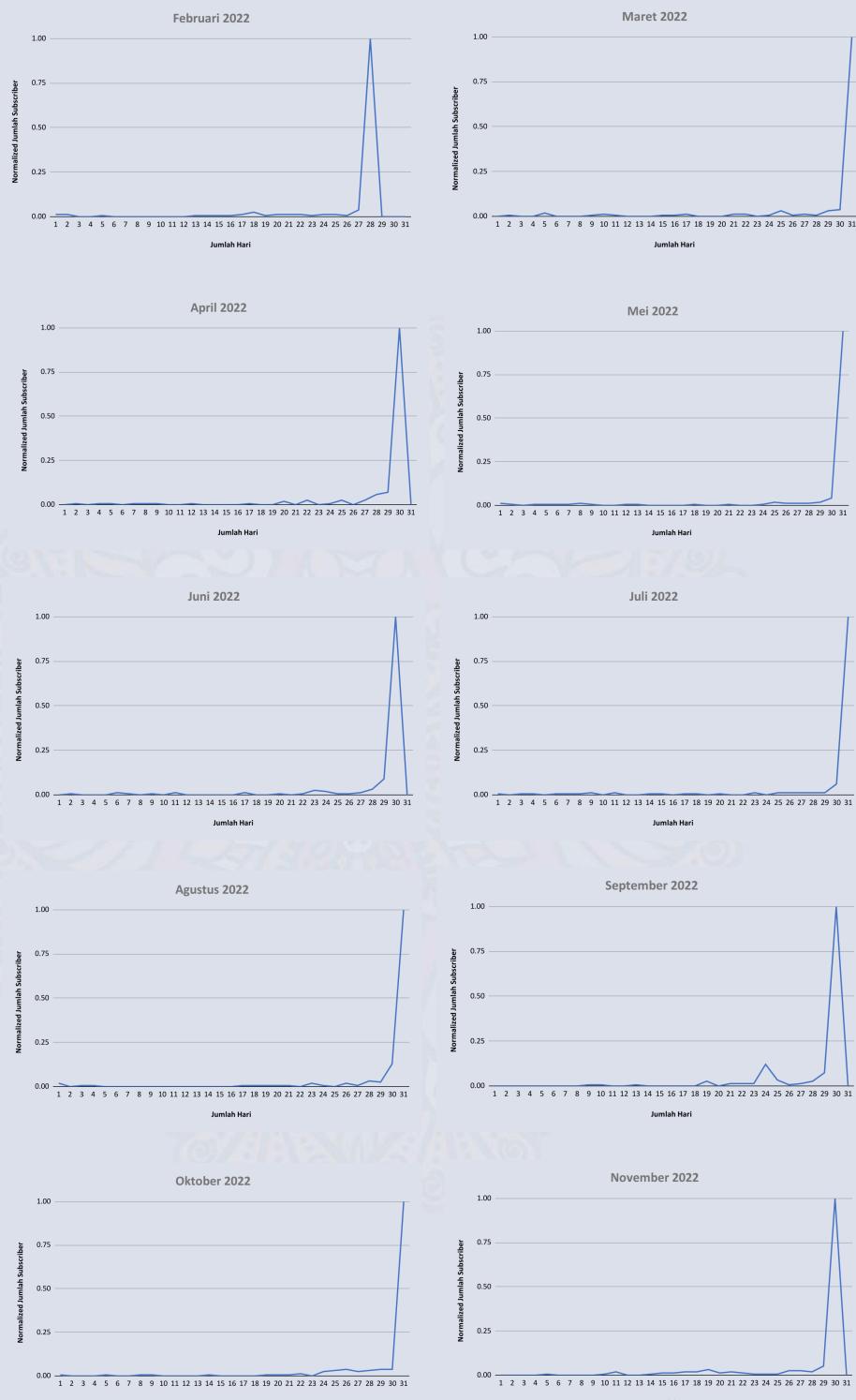


Gambar 4.31 Jumlah Subscribers per Hari Data Testing Setelah Records Duplicate Dihapus

Dari grafik diatas terlihat bahwa jumlah *subscribers* terbanyak pada data *testing* adalah *subscribers* dengan *records* bertipe *data source* LBA sedangkan jumlah *subscribers* paling sedikit adalah *subscribers* dengan *records* bertipe *data source* CHG. Namun, berbeda dengan *data training*, pada grafik terlihat bahwa jumlah *records* untuk ketiga *data source* relatif konstan, meskipun terdapat *spike* pada jumlah *subscribers* yang memiliki *records* dengan tipe *data source* UPCC pada beberapa hari di Bulan Januari.

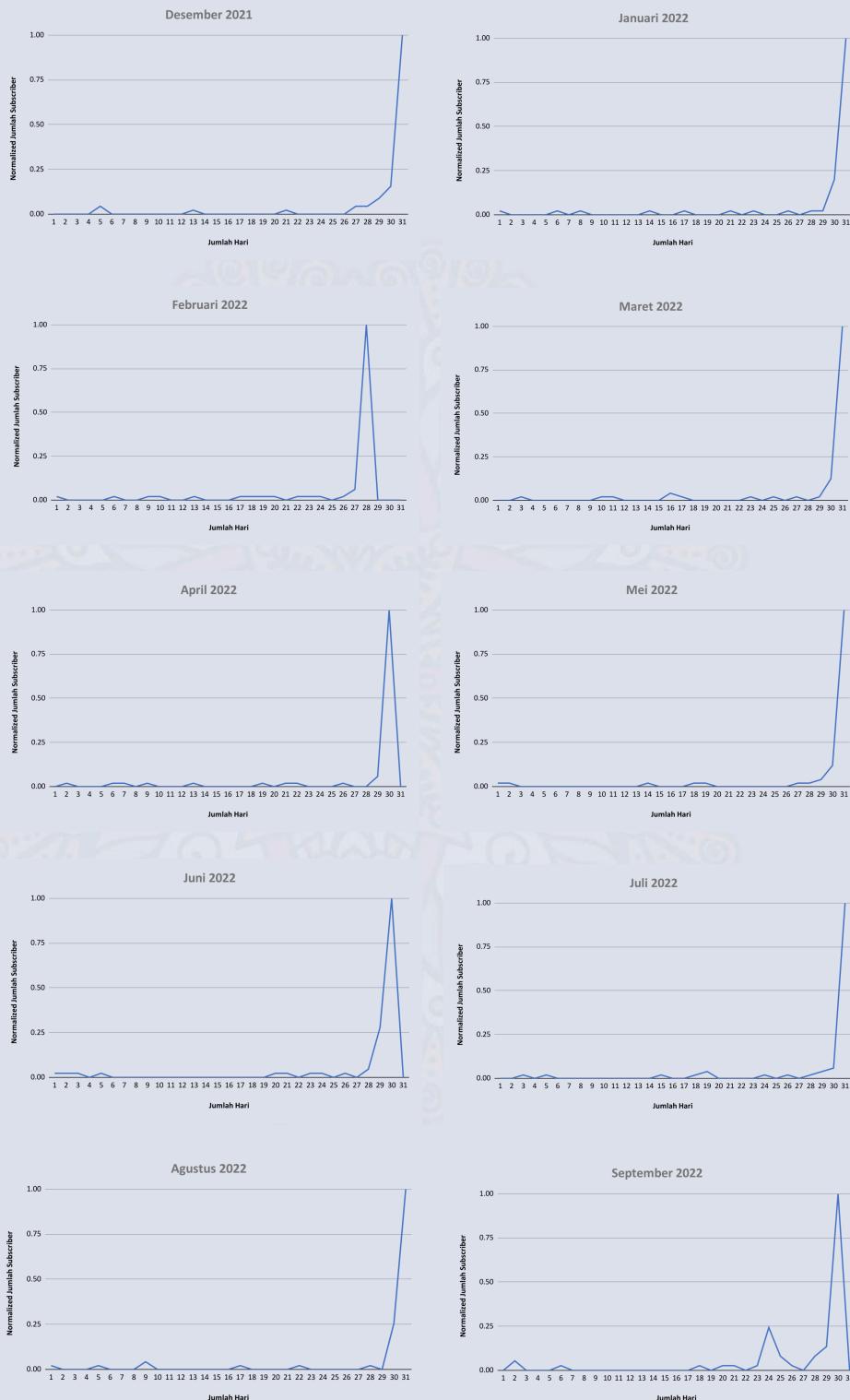
Jumlah Subscribers Data Training yang Hadir per Bulan

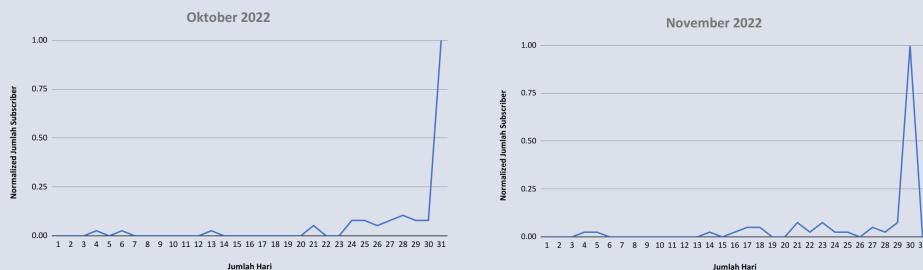




Gambar 4.32 Jumlah Subscribers Data *Training* yang Hadir per Bulan

Jumlah Subscribers Data Testing yang Hadir per Bulan





Gambar 4.33 Jumlah Subscribers Data *Testing* yang Hadir per Bulan

Berdasarkan gambar di atas terlihat bahwa pada periode Desember 2021 hingga November 2022, secara keseluruhan, normalisasi jumlah pengguna untuk data training dan testing yang hadir masing-masing mencapai angka 1 pada hari-hari terakhir di setiap bulannya. Hal tersebut membuktikan bahwa mayoritas pengguna hadir secara lengkap atau menyeluruh pada satu bulan tersebut. Akan tetapi, masih terdapat beberapa *spike* yang sangat kecil pada hari-hari tertentu yang menunjukkan bahwa masih terdapat beberapa pengguna yang tidak hadir secara lengkap atau menyeluruh di bulan tersebut.

Secara keseluruhan, dapat disimpulkan bahwa kualitas *data training* dan *data testing* sudah cukup baik untuk digunakan dalam pemrosesan selanjutnya. Namun, kualitas *data training* sedikit lebih baik dibandingkan dengan kualitas *data testing* jika dilihat berdasarkan distribusi jumlah *records*.

4.4 Pembangunan Algoritma *Family Grouping* dan *Double Counting*

Data *training* yang telah dikontrol kualitasnya kemudian digunakan untuk membangun algoritma *family grouping* dan *double counting*. Tahap awal dalam pembangunan kedua algoritma tersebut adalah menentukan lokasi *usual environment* setiap *subscribers*. Lokasi ini kemudian digunakan untuk mengelompokkan *subscribers* yang memiliki *usual environment* yang mirip satu sama lain. Teridentifikasinya *subscribers* yang memiliki kemiripan *usual environment* akan

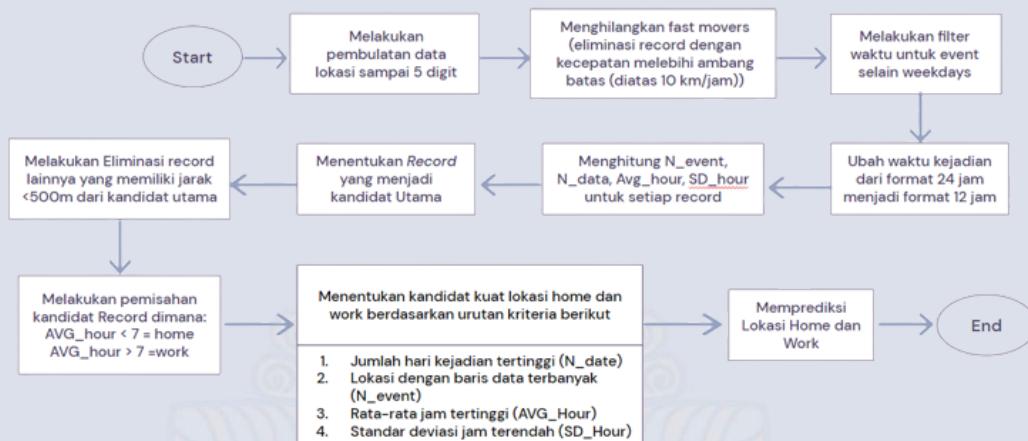
memperbesar peluang bahwa *subscribers* tersebut merupakan entitas *family grouping* atau *double counting*.

4.4.1 Penentuan *Usual Environment* Menggunakan AMDA

Suatu nomor yang digunakan untuk aktivitas seluler seperti telepon, SMS, dan internet akan secara otomatis menghasilkan data waktu dan lokasi yang disebut sebagai *passive mobile positioning data*. Data tersebut berisi waktu (tanggal dan jam), tipe *data source*, serta *latitude* dan *longitude*, yang selanjutnya dapat digunakan untuk mengidentifikasi lokasi dalam lingkup provinsi, kabupaten/kota, dan kecamatan. Titik-titik lokasi yang dihasilkan pada waktu-waktu tertentu bisa saja memiliki kesamaan satu sama lain atau dengan kata lain, lokasi nomor masih berada di satu daerah yang sama dengan lokasi yang tercatat sebelumnya. Kecenderungan lokasi-lokasi yang sama dalam suatu waktu tertentu akan membentuk sebuah pola sehingga dapat dijadikan sebagai informasi untuk menentukan tempat dimana nomor tersebut sering berada atau disebut dengan *usual environment*.

BPS (2021) mendefinisikan *usual environment* sebagai wilayah atau area lokasi lingkungan keseharian pengguna seluler yang terdiri dari wilayah rumah (*home*), lokasi bekerja (*work*), serta rumah kedua pengguna seluler. Lokasi rumah (*home*) merupakan tempat penting yang menjadi langkah awal untuk penelusuran mobilitas manusia lebih lanjut (Bojic et al., 2015). Hal ini kemudian juga berlaku untuk lokasi kerja (*work*) yang dapat menjadi dasar penentuan analisis perencanaan pergerakan untuk komuter, transportasi, dan lalu lintas (Putra et al., 2021). Dengan demikian, pemilihan *usual environment* berupa lokasi rumah (*home*) dan kerja (*work*) menjadi penting untuk dilakukan. Untuk menentukan *usual environment* setiap *subscribers*, penelitian ini menggunakan metode Anchor Mobility Data Analytic (AMDA). Tahapan-tahapan yang dilakukan untuk menentukan *usual environment* dengan AMDA terlihat pada gambar 4.34.

Flowchart Tahapan Penentuan *Usual Environment* dengan AMDA



Gambar 4.34 Flowchart Tahapan Penentuan *Usual Environment* dengan AMDA

Dalam AMDA, proses untuk menentukan *usual environment* dimulai dari Master Raw. Proses ini bertujuan untuk mengidentifikasi anomali pada setiap perpindahan lokasi yang terjadi. Terdapat dua tahapan utama dalam Master Raw (Putra et al., 2021), yaitu:

1. Pembulatan lokasi. Tahapan ini dilakukan dengan membulatkan koordinat *latitude* dan *longitude* dengan ketelitian lima desimal. Langkah ini bertujuan untuk menggabungkan titik lokasi *cell-id* yang sangat berdekatan. Contoh perhitungannya adalah sebagai berikut.

Tabel 4.4 Contoh Record MPD Sebelum Dilakukan Pembulatan

msisdn	datetime	source	latitude	longitude
msisdn00xxx	2021-01-01 12:36:41	LBA	-7.795097	110.358727
msisdn00xxx	2021-01-01 13:31:37	LBA	-7.804648	110.357843
msisdn00xxx	2021-01-01 14:25:37	LBA	-7.804648	110.357843

msisdn00xxx	2021-01-01 14:42:24	LBA	-7.76936	110.364236
-------------	---------------------	-----	----------	------------

msisdn00xxx	2021-01-01 16:30:51	LBA	-7.783124	110.398203
-------------	---------------------	-----	-----------	------------

Tabel 4.5 Contoh Record MPD Setelah Dilakukan Pembulatan

msisdn	datetime	source	latitude	longitude
msisdn00xxx	2021-01-01 12:36:41	LBA	-7.79510	110.35873
msisdn00xxx	2021-01-01 13:31:37	LBA	-7.80465	110.35784
msisdn00xxx	2021-01-01 14:25:37	LBA	-7.80465	110.35784
msisdn00xxx	2021-01-01 14:42:24	LBA	-7.76940	110.36424
msisdn00xxx	2021-01-01 16:30:51	LBA	-7.78312	110.39820

2. Menghilangkan *fast movers*. Tahapan ini bertujuan untuk menghilangkan perpindahan yang tidak lazim antara dua titik. Hal tersebut dilakukan dengan menghapus *record* dengan kecepatan melebihi ambang batas (diatas 10 km/jam). Proses ini dilakukan dengan menghitung jarak dari urutan setiap lokasi MSISDN menggunakan formula Haversine. Berikut adalah contoh perhitungan jarak Haversine untuk dua koordinat awal pada tabel yang sudah dilakukan pembulatan, yaitu koordinat (-7.79510, 110.35873) dan (-7.80465, 110.35784).

Diketahui:

$$\varphi_1 = -7.7951 = -0.1361 \text{ rad}$$

$$\varphi_2 = -7.80465 = -0.1362 \text{ rad}$$



$$\lambda_1 = 110.35873 = 1.926 \text{ rad}$$

$$\lambda_2 = 110.35784 = 1.926 \text{ rad}$$

$$r = 6,371 \text{ km (radius bumi)}$$

$$d = 2r \times \arcsin \left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)} \right)$$

$$d = 2 \times (6,371) \times$$

$$\arcsin \left(\sqrt{\sin^2\left(\frac{-0.1362 - (-0.1361)}{2}\right) + \cos(-0.1361) \cdot \cos(-0.1362) \cdot \sin^2\left(\frac{1.926 - 1.926}{2}\right)} \right)$$

$$d = 2 \times (6,371) \times$$

$$\arcsin \left(\sqrt{\sin^2(-0.00005) + \cos(-0.1361) \cdot \cos(-0.1362) \cdot \sin^2(0)} \right)$$

$$d = 2 \times (6,371) \times \arcsin(9.009119471733968e - 10)$$

$$d = 0,382454$$

Dari hasil penghitungan di atas, diperoleh nilai jarak Haversine antara dua koordinat tersebut adalah sebesar 0,382454. Hasil perhitungan jarak Haversine dengan menggunakan koordinat yang berbeda dapat dilihat di tabel 4.6.

Tabel 4.6 Contoh Record MPD dengan Penghitungan
Haversine Distance

msisdn	datetime	source	latitude	longitude	haversine distance
msisdn00xxx	1/1/2021 12:36	LBA	-7.79510	110.35873	
msisdn00xxx	1/1/2021 13:31	LBA	-7.80465	110.35784	0.382454
msisdn00xxx	1/1/2021 14:25	LBA	-7.80465	110.35784	0

msisdn00xxx	1/1/2021	LBA	-7.76940	110.36424	1.538283
			14:42		
msisdn00xxx	1/1/2021	LBA	-7.78312	110.3982	3.813374
			16:30		

Setelah diketahui jarak dari dua pasang koordinat, selanjutnya dilakukan penghitungan waktu yang dibutuhkan untuk berpindah dari satu koordinat ke koordinat lainnya secara berurutan. Variabel waktu ini diperlukan di dalam penghitungan untuk mendapatkan variabel kecepatan. Berikut adalah tabel contoh *record* data yang telah dilakukan perhitungan kecepatan antar titik koordinatnya.

Tabel 4.7 Contoh *Record* MPD dengan Penghitungan Kecepatan

msisdn	datetime	source	latitude	longitude	Haversine distance	Time (jam)	Speed (km/jam)
msisdn 00xxx	1/1/2021 12:36	LBA	-7.79510	110.35873			
msisdn 00xxx	1/1/2021 13:31	LBA	-7.80465	110.35784	0.382454	0.916667	0.417223
msisdn 00xxx	1/1/2021 14:25	LBA	-7.80465	110.35784		0	0.9
msisdn 00xxx	1/1/2021 14:42	LBA	-7.76940	110.36424	1.538283	0.283333	5.429234
msisdn 00xxx	1/1/2021 16:30	LBA	-7.78312	110.3982	3.813374	1.8	2.118541



Berdasarkan Tabel 4.7, terlihat bahwa tidak ada kecepatan yang melebihi 10 km/jam. Oleh karena itu, tidak ada *record* yang dieliminasi. Setelah itu, dilakukan pendekripsi lokasi rumah (*home*) menggunakan AMDA dengan membentuk *anchor point* melalui tahapan berikut (Putra et al., 2021):

1. Untuk setiap nomor/*subscriber*, filter *event* pertama yang terekam setiap harinya pada hari kerja (*weekdays*).
2. Ubah waktu kejadian yang masih menggunakan format 24 jam menjadi format 12 jam dengan mengikuti algoritma sebagai berikut.
 - a. Melakukan pengecekan untuk setiap *timestamp* apakah jam yang tertera lebih besar dari 12 atau tidak.
 - b. Jika diperoleh hasilnya “Ya” ($\text{Jam} > 12$), maka lakukan transformasi pada *Timestamp* dengan menggunakan formula: $\text{jam} - 12$.
 - c. Jika diperoleh hasilnya “Tidak”, maka tidak akan dilakukan transformasi pada *Timestamp*.
 - d. Selanjutnya, untuk setiap *subscriber*, hitung frekuensi berada di LAU tertentu (dalam hal ini LAU2 Regency)

Format 24 jam (*Original*):

0	1	2	3	4	5	6	7	8	9	10	11
23	22	21	20	19	18	17	16	15	14	13	12

Format 12 jam (*Converted*):

0	1	2	3	4	5	6	7	8	9	10	11
11	10	9	8	7	6	5	4	3	2	1	0

3. Hitung frekuensi kemunculan awal *event* per hari per cell-id (*anchor*).
4. Hitung statistik rata-rata dan standar deviasi untuk jam terjadinya *event*, jumlah hari, dan jumlah kemunculan *event*. Perhitungan ini dilakukan untuk setiap *anchor*.

Setelah menentukan *anchor* point, langkah selanjutnya adalah menentukan kandidat *usual environment* melalui tahapan berikut (Putra et al., 2021) :

1. Hitung jarak antara *anchor* dan kandidat utama. Penghitungan jarak dilakukan menggunakan *haversine distance*.
2. Filter *anchor* dengan jumlah hari yang muncul lebih dari atau sama dengan 5 (≥ 5). Eliminasi *anchor* yang berdekatan dengan kandidat utama (kurang dari 500 m).
3. Pisahkan kejadian (*event*) yang ada hanya pada hari kerja dan ubah variabel waktu menjadi format 12 jam.
4. Hitung frekuensi lokasi awal harian dari setiap titik lokasi (*anchor*), diikuti dengan statistik rata-rata dan standar deviasi setiap jam kejadian, dan bentuk kandidat utamanya.
5. Hitung jarak antara lokasi yang dipilih dengan lokasi kandidat utama dan kemudian eliminasi lokasi yang berdekatan (kurang dari 500 m) diikuti dengan mengurutkan lama kejadian yang lebih atau sama dengan 5 hari (≥ 5 hari).

Secara garis besar, terdapat proses pembentukan dua *anchor* dalam AMDA. Pada *anchor* 1, *filtering* dilakukan hanya untuk waktu hari kerja pada data. Jam kejadian kemudian diubah menjadi format 12 jam menggunakan algoritma yang telah dijelaskan sebelumnya. Setelah itu, dilakukan penghitungan jumlah *record* (*N_event*), jumlah hari unik (*N_date*), serta rata-rata jam kejadian (*AVG_hour*) diikuti dengan standar deviasi jam kejadian (*SD_hour*) untuk setiap nomor/*subscriber* per bulannya. Semua fitur ini nantinya akan digunakan untuk menentukan



usual environment. Secara khusus, standar deviasi jam kejadian akan digunakan sebagai *filter* terakhir ketika data memiliki jumlah hari unik, jumlah baris data (*record*), dan rata-rata jam kejadian yang sama.

Tabel 4.8 Contoh Agregasi Penghitungan *Anchor* 1

msisdn	year	month	latitude	longitude	idkab	N_event	N_date	Avg_hour	SD_hour
msisdn 00xxx	2022	1	-6.9201	107.5607	32 77	113	23	1.05	2.07
msisdn 00xxx	2022	1	-6.922	107.5559	32 77	46	22	3.83	3.03
msisdn 00xxx	2022	1	-6.9201	107.5607	32 77	155	19	3.6	3.55
msisdn 00xxx	2022	1	-6.922	107.5559	32 77	67	18	4.31	4.05
msisdn 00xxx	2022	1	-7.0282	107.5228	32 04	105	12	8.89	0.8

Selanjutnya, pada pembentukan *anchor* 2, dilakukan pengecekan jumlah kemunculan suatu lokasi dengan pendekatan minimal setidaknya muncul sebanyak lima hari setiap bulan. Tahapan ini melihat potensi dari kandidat utama dengan melakukan pengelompokan/*clustering* setiap lokasi yang ada. Pengelompokan dilakukan dengan menghitung jarak antar lokasi yang dihitung menggunakan jarak Haversine, kemudian dilakukan pengelompokan lokasi-lokasi dengan radius jarak 0,5 km. Lokasi yang dianggap dekat dengan kandidat utama pada akhirnya dihilangkan dari calon lokasi *usual environment*.

Tabel 4.9 Contoh Agregasi Penghitungan *Anchor* 2

msisdn	year	month	latitude	longitude	idkab	N_event	N_date	AVG_hour	SD_hour	distance
msisdn 00xxx	2022	1	-6.9201	107.5607	32 77	113	23	1.05	2.07	0
msisdn 00xxx	2022	1	-6.922	107.5559	32 77	46	22	3.83	3.03	0.2
msisdn 00xxx	2022	1	-6.9201	107.5607	32 77	155	19	3.6	3.55	0.76
msisdn 00xxx	2022	1	-6.922	107.5559	32 77	67	18	4.31	4.05	17.22
msisdn 00xxx	2022	1	-7.0282	107.5228	32 04	105	12	8.89	0.8	20.15

Baris yang berwarna hijau pada Tabel 4.9 adalah lokasi yang menjadi kandidat utama karena memiliki jumlah hari unik (*N_date*) terbanyak dibandingkan dengan lokasi lainnya. Selanjutnya, setiap lokasi terhadap lokasi yang menjadi kandidat utama. Baris berwarna merah merupakan lokasi yang dieliminasi dari calon lokasi *usual environment* karena berjarak kurang dari 0,5 km dari lokasi yang menjadi kandidat utama. Sementara tiga lokasi lainnya tidak akan dieliminasi karena berjarak lebih dari 0,5 km dari lokasi yang menjadi kandidat utama.

Usual environment didefinisikan menjadi dua lokasi, yaitu lokasi rumah (*home*) dan kantor (*work*). *Anchor* hasil pengelompokan lokasi dengan rata-rata jam kurang dari tujuh (*AVG_hour* < tujuh) akan diklasifikasikan sebagai calon lokasi *home*, sedangkan *anchor* lainnya dengan rata-rata jam lebih dari tujuh akan diklasifikasikan sebagai calon lokasi *work*. Kemudian, pilih kandidat terkuat untuk memprediksi lokasi



home dan *work* dari setiap kelompok dengan urutan prioritas kriteria sebagai berikut:

1. jumlah hari kejadian tertinggi (N_date)
2. lokasi dengan baris data terbanyak (N_event)
3. rata-rata jam tertinggi (AVG_hour), dan
4. standar deviasi jam terendah (SD_hour)

Hasil akhir dari metode AMDA ini berupa prediksi lokasi *home* dan *work* yang ditunjukkan pada contoh tabel berikut.

Tabel 4.10 Prediksi Lokasi *Usual Environment*

msisdn	year	month	latitude	longitude	idkab	N_event	N_date	AVG_hour	SD_hour	label
msisdn 00xxx	2020	1	-6.2500	106.9067	31 72	39	22	1.07	1.96	Home
msisdn 00xxx	2020	1	-7.0282	107.5228	32 04	105	12	8.89	0.8	Work

4.4.2 Pengelompokan *Subscribers* dengan Lokasi *Home* yang Sama

Setelah ditentukan *usual environment* untuk setiap *subscribers* berupa lokasi *home* dan *work*, selanjutnya *subscribers* yang memiliki lokasi *home* yang sama akan dikelompokkan menjadi satu kelompok. Pengelompokan ini bertujuan untuk mengurangi beban komputasi pengolahan data karena nomor-nomor milik *subscribers* yang berada pada kelompok yang sama akan dibandingkan, sedangkan nomor-nomor yang berada pada kelompok berbeda tidak akan dibandingkan. Lokasi *home* yang digunakan untuk pengelompokan adalah lokasi *home* tahunan (*yearly home*) setingkat kabupaten/kota dan tidak mengikutsertakan lokasi *work* dari *subscribers*. Hal ini didasarkan pada keakuratan (akurasi) metode

AMDA yang lebih baik dalam mengidentifikasi *home* pelanggan (provinsi: 98,8%, kabupaten/kota: 88,7%) daripada mengidentifikasi *work* pelanggan (provinsi: 98,9%, kabupaten/kota: 70,4%) (Putra et al., 2021).

Lokasi dengan jumlah kemunculan tertinggi (modus) di setiap bulan akan menjadi kandidat kuat untuk menjadi *usual environment* (lokasi *home* dan *work*) (Putra et al., 2021). Berdasarkan hal tersebut, untuk menentukan lokasi *home* tahunan, akan diambil modus dari lokasi *home* nomor-nomor yang telah dilakukan penentuan lokasi *home* perbulan. Modus yang diambil merupakan modus lokasi *home* selama rentang waktu 12 bulan ke belakang, yaitu dari bulan Desember 2021 sampai bulan November 2022. Jika terdapat dua atau lebih modus *home* dengan lokasi yang berbeda, maka lokasi *home* ditentukan, secara berurutan, berdasarkan lokasi dengan jumlah hari kejadian tertinggi (N_date), lokasi dengan baris data terbanyak (N_event), lokasi dengan rata-rata jam tertinggi (AVG_hour), dan lokasi dengan standar deviasi jam terendah (SD_hour) diantara bulan-bulan yang menjadi modus lokasi *home* (Putra et al., 2021). Setelah didapatkan lokasi *home* untuk setiap *subscribers/nomor*, selanjutnya nomor-nomor yang memiliki lokasi *home* pada provinsi dan kabupaten/kota yang sama akan berada dalam satu kelompok, seperti yang dicontohkan pada tabel berikut.

Tabel 4.11 Contoh Pengelompokan *Subscribers* dengan Lokasi Home yang Sama

msisdn	home_provinsi	home_kabupaten	provider	kelompok
msisdn00xxx31	DKI JAKARTA	JAKARTA UTARA	Telkomsel	
msisdn00xxx87	DKI JAKARTA	JAKARTA UTARA	Telkomsel	1
msisdn00xxx5	JAWA BARAT	BEKASI	Telkomsel	



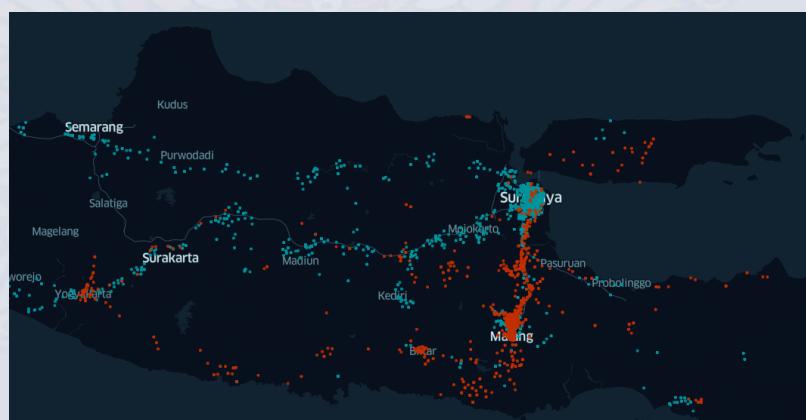
msisdn00xxx6	JAWA BARAT	BEKASI	Telkomsel	
msisdn00xxx7	JAWA BARAT	BEKASI	Telkomsel	2
msisdn00xxx8	JAWA BARAT	BEKASI	Telkomsel	
msisdn00xxx19	JAWA BARAT	CIMAHI	Telkomsel	3
msisdn00xxx1	JAWA BARAT	DEPOK	Telkomsel	
msisdn00xxx3	JAWA BARAT	DEPOK	Telkomsel	
msisdn00xxx67	JAWA BARAT	DEPOK	Telkomsel	4
msisdn00xxx94	JAWA BARAT	DEPOK	Telkomsel	
msisdn00xxx105	JAWA BARAT	SUMEDANG	Telkomsel	5
msisdn00xxx10	JAWA TENGAH	BOYOLALI	Telkomsel	
msisdn00xxx108	JAWA TENGAH	BOYOLALI	Telkomsel	6
msisdn00xxx27	JAWA TENGAH	JEPARA	Telkomsel	
msisdn00xxx104	JAWA TENGAH	JEPARA	Indosat	7

4.4.3 Eksplorasi Temuan

4.4.3.1 Eksplorasi Kasus Wilayah yang Berbatasan Memiliki *Home* Berbeda

Pengelompokan *subscribers* yang memiliki *home* yang sama dilakukan dengan algoritma yang telah dijelaskan sebelumnya dan dengan hasil yang dapat terlihat pada tabel 4.11. Namun, setelah dilakukan pengelompokan, masih terdapat beberapa kasus temuan yang menarik. Salah satunya adalah kasus ketika *subscribers* diprediksi memiliki lokasi *home* yang berbeda pada wilayah yang berbatasan secara geografis. Seperti yang ditunjukkan pada salah satu kasus *subscriber* yang memiliki dua nomor dengan id msisdn00260 dan msisdn00325 yang digunakan pada dua ponsel utamanya.

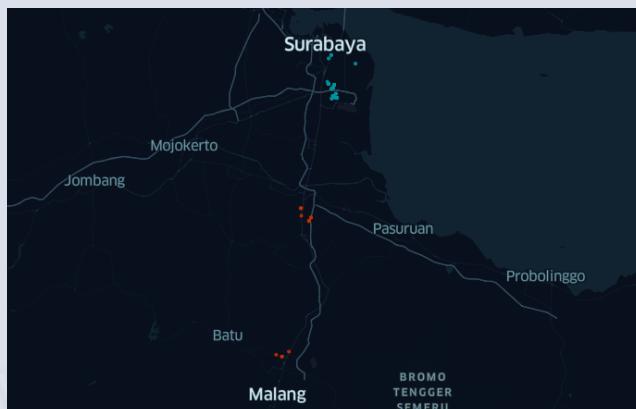
Sebaran *Raw Data* msisdn00260 (Hijau) dan msisdn00325 (Merah)



Gambar 4.35 Sebaran *Raw Data* msisdn00260 (Hijau) dan msisdn00325 (Merah)

Berdasarkan visualisasi di atas, terlihat bahwa sebaran *raw data* berupa titik-titik *records* milik msisdn00260 dan msisdn00325 terkonsentrasi pada dua wilayah, yaitu wilayah di sekitar Surabaya dan Malang. Untuk menentukan lokasi *home* masing-masing nomor, diimplementasikan algoritma AMDA sehingga diperoleh kandidat lokasi *home* seperti yang ditunjukkan pada visualisasi di bawah ini.

**Hasil Penentuan Kandidat Lokasi *Home* msisdn00260 (Hijau) dan
msisdn00325 (Merah) Menggunakan AMDA**



Gambar 4.36 Hasil Penentuan Kandidat Lokasi *Home* msisdn00260 (Hijau) dan msisdn00325 (Merah) Menggunakan AMDA

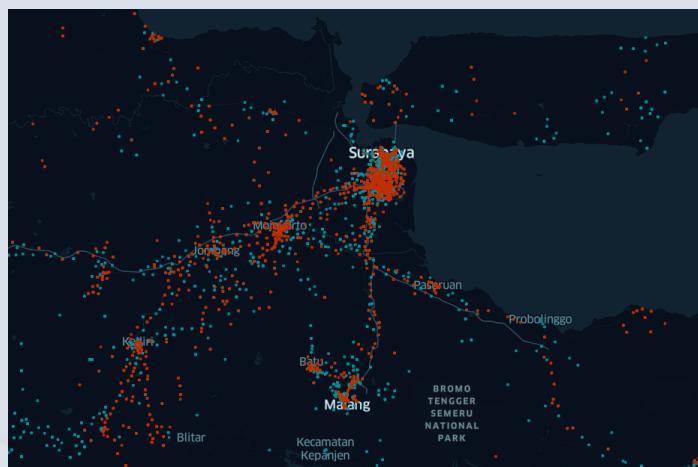
Berdasarkan visualisasi di atas, terlihat bahwa msisdn00260 memiliki kandidat lokasi *home* di Kota Surabaya dan Kabupaten Sidoarjo, sedangkan msisdn00260 memiliki kandidat lokasi *home* di Kabupaten Pasuruan dan Kota Malang. Hasil akhir prediksi lokasi *home* dari kandidat lokasi *home* tersebut terlihat pada tabel di bawah ini.

Tabel 4.12 Prediksi Lokasi *Home* msisdn00260 dan msisdn00325 dengan AMDA

msisdn	home_provinsi	home_kabupaten	provider
msisdn00260	JAWA TIMUR	KABUPATEN SIDOARJO	Telkomsel
msisdn00325	JAWA TIMUR	KABUPATEN PASURUAN	Telkomsel

Berdasarkan Tabel 4.12, dapat disimpulkan bahwa *subscribers* tersebut memiliki lokasi *home* yang berbeda untuk masing-masing nomor. msisdn00260 diprediksi memiliki lokasi *home* di Kabupaten Sidoarjo, sedangkan msisdn00260 diprediksi memiliki lokasi *home* di Kabupaten Pasuruan yang secara geografis berbatasan dengan Kabupaten Sidoarjo. Selain itu, contoh kasus lokasi *home* yang berbeda pada wilayah berbatasan juga ditemukan pada relawan yang memiliki dua nomor dengan id msisdn00288 dan msisdn00331 yang digunakan pada satu *handphone* utama.

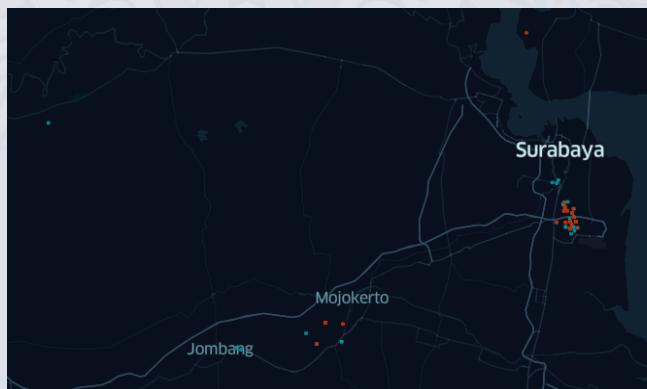
Sebaran *Raw Data* msisdn00288 (Hijau) dan msisdn00331 (Merah)



Gambar 4.37 Sebaran *Raw Data* msisdn00288 (Hijau) dan msisdn00331 (Merah)

Berdasarkan visualisasi di atas, terlihat bahwa sebaran *raw data* berupa titik-titik *records* milik msisdn00288 dan msisdn00331 terkonsentrasi pada dua wilayah, yaitu wilayah di sekitar Surabaya dan Mojokerto. Setelah diimplementasikan algoritma AMDA, diperoleh kandidat lokasi *home* untuk masing-masing nomor seperti pada visualisasi di bawah ini.

Hasil Penentuan Kandidat Lokasi Home msisdn00288 (Hijau) dan msisdn00331 (Merah) Menggunakan AMDA



Gambar 4.38 Hasil Penentuan Kandidat Lokasi Home msisdn00288 (Hijau) dan msisdn00331 (Merah) Menggunakan AMDA

Berdasarkan visualisasi di atas, terlihat bahwa baik msisdn00288 maupun msisdn00331 memiliki kandidat lokasi *home* di Kota Surabaya,



Kabupaten Sidoarjo, dan Kabupaten Mojokerto. Hasil akhir prediksi lokasi *home* dari kandidat lokasi *home* tersebut terlihat pada tabel di bawah ini.

Tabel 4.13 Prediksi Lokasi *Home* msisdn00288 dan msisdn00331 dengan AMDA

msisdn	home_provinsi	home_kabupaten	provider
msisdn00288	JAWA TIMUR	KABUPATEN SIDOARJO	Telkomsel
msisdn00331	JAWA TIMUR	KABUPATEN MOJOKERTO	Telkomsel

Berdasarkan Tabel 4.13, dapat disimpulkan bahwa *subscribers* tersebut memiliki lokasi *home* yang berbeda untuk masing-masing nomor. msisdn00288 diprediksi memiliki lokasi *home* di Kabupaten Sidoarjo, sedangkan msisdn00331 diprediksi memiliki lokasi *home* di Kabupaten Mojokerto yang secara geografis berbatasan dengan Kabupaten Sidoarjo.

Kedua kasus di atas dapat menyebabkan kesalahan dalam pengelompokan *subscribers* berdasarkan lokasi *home* yang sama. Seharusnya dua nomor yang digunakan dalam satu *handphone* utama milik setiap *subscribers* memiliki lokasi *home* yang sama dan dikelompokkan menjadi satu kelompok. Akan tetapi, kedua nomor tersebut dikelompokkan ke dalam kelompok yang berbeda karena lokasi *home* yang berbeda.

4.4.3.2 Eksplorasi Kasus Nomor yang Digunakan di Waktu yang Berbeda Memiliki *Home* Berbeda

Selain kasus *home* yang berbeda karena wilayah yang berbatasan, ditemukan juga kasus *home* yang berbeda karena nomor-nomor yang digunakan di waktu yang berbeda. Kasus ini ditemukan terjadi karena terdapat beberapa relawan Politeknik Statistika STIS yang membeli nomor baru ketika berpindah lokasi tinggal dari rumahnya di daerah masing-masing menjadi di Jakarta Timur untuk melaksanakan perkuliahan secara luring. Seperti yang ditunjukkan pada salah satu kasus *subscribers*

yang memiliki dua nomor dengan id msisdn00019 dan msisdn00096 yang digunakan pada satu *handphone* utama.

Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah) Selama Rentang

Waktu 1 Desember 2021 - 22 Oktober 2022

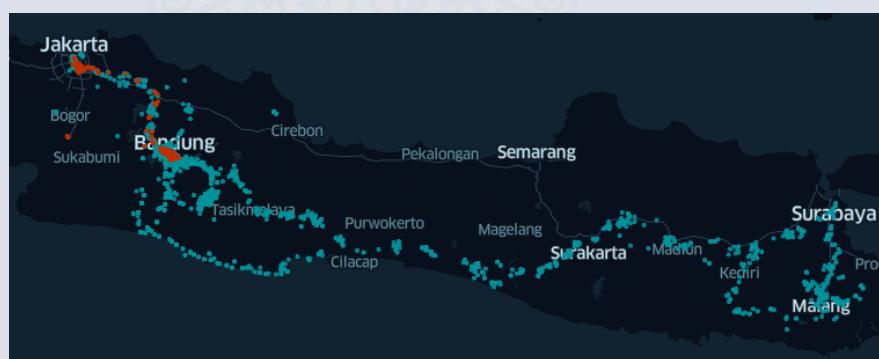


Gambar 4.39 Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah)
Rentang Waktu 1 Desember 2021 - 22 Oktober 2022

Berdasarkan visualisasi di atas, terlihat bahwa selama rentang waktu 1 Desember 2021 - 22 Oktober 2022 hanya terdapat sebaran *raw data* milik msisdn00019 dan tidak ada *raw data* milik msisdn00096. Hal ini menunjukkan bahwa selama rentang waktu tersebut, *subscribers* hanya menggunakan satu nomor/kartu sim yaitu msisdn00019. Selain itu, terlihat juga bahwa sebaran *raw data* berupa titik-titik records milik msisdn00019 terkonsentrasi di wilayah Bandung dan Cimahi Provinsi Jawa Barat.

Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah) Selama Rentang

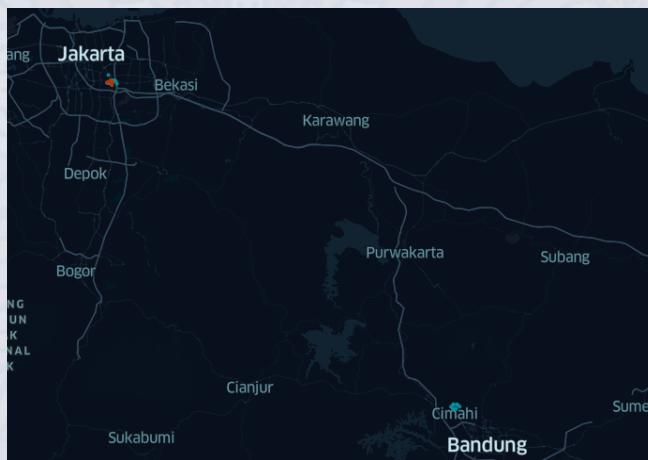
Waktu 1 Desember 2021 - 30 November 2022



Gambar 4.40 Sebaran Raw Data msisdn00019 (Hijau) dan msisdn00096 (Merah)
Rentang Waktu 1 Desember 2021 - 30 November 2022

Jika rentang waktu diperpanjang hingga 30 November 2022, maka mulai muncul sebaran *raw data* milik msisdn00096. Hal ini menunjukkan bahwa *subscribers* mulai menggunakan nomor/kartu sim msisdn00096 sejak tanggal 23 Oktober 2022 yang berakibat pada berbedanya periode waktu penggunaan antara msisdn00019 dan msisdn00096. Terlihat juga bahwa sebaran *raw data* berupa titik-titik records milik msisdn00096 cenderung terkonsentrasi di wilayah DKI Jakarta, khususnya Kota Jakarta Timur. Selanjutnya, diimplementasikan algoritma AMDA untuk menentukan lokasi *home* masing-masing nomor sehingga diperoleh kandidat lokasi *home* seperti yang ditunjukkan pada visualisasi di bawah ini.

Hasil Penentuan Kandidat Lokasi *Home* msisdn00019 (Hijau) dan msisdn00096 (Merah) Menggunakan AMDA



Gambar 4.41 Hasil Penentuan Kandidat Lokasi *Home* msisdn00019 (Hijau) dan msisdn00096 (Merah) Menggunakan AMDA

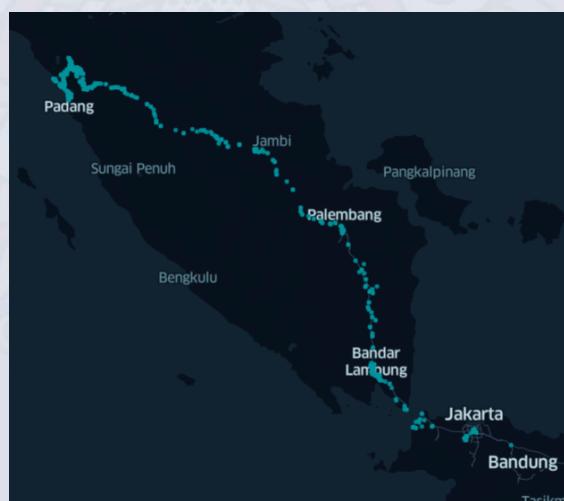
Dari visualisasi di atas, terlihat bahwa msisdn00019 memiliki kandidat lokasi *home* di Kota Cimahi dan Kota Jakarta Timur, sedangkan msisdn00096 hanya memiliki kandidat lokasi *home* di Kota Jakarta Timur. Hasil akhir prediksi lokasi *home* dari kandidat lokasi *home* tersebut adalah sebagai berikut.

Tabel 4.14. Prediksi Lokasi *Home* msisdn00019 dan msisdn00096 dengan AMDA

msisdn	home_provinsi	home_kabupaten	provider
msisdn00019	JAWA BARAT	KOTA CIMAHI	Telkomsel
msisdn00096	DKI JAKARTA	KOTA JAKARTA TIMUR	Telkomsel

Berdasarkan Tabel 4.14, dapat diketahui bahwa *subscribers* tersebut memiliki lokasi *home* yang berbeda untuk masing-masing nomor. msisdn00019 diprediksi memiliki lokasi home di Kota Cimahi, Jawa Barat sedangkan msisdn00096 diprediksi memiliki lokasi home di Kota Jakarta Timur, DKI Jakarta. Contoh kasus lokasi *home* yang berbeda karena nomor yang digunakan pada periode waktu berbeda juga ditemukan pada *subscribers* yang memiliki dua nomor dengan id msisdn00079 dan msisdn00080 yang digunakan pada satu *handphone* utama.

Sebaran Raw Data msisdn00079 (Hijau) dan msisdn00080 (Merah) Selama Rentang Waktu 1 Desember 2021 - 27 Juli 2022

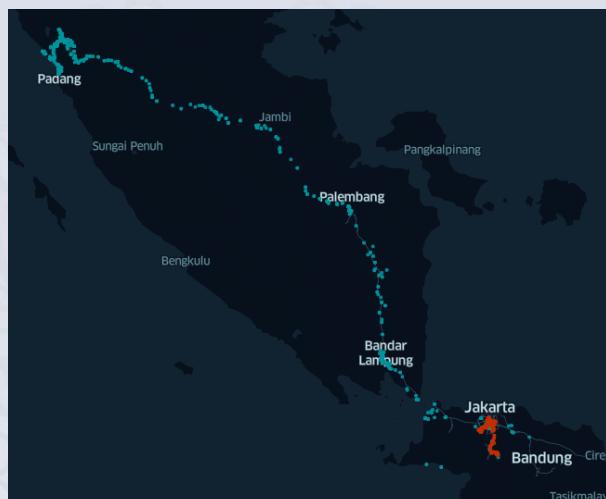


Gambar 4.42 Sebaran *Raw Data* msisdn00079 (Hijau) dan msisdn00080 (Merah) Rentang Waktu 01 Desember 2021 - 27 Juli 2022

Berdasarkan visualisasi di atas, terlihat bahwa selama rentang waktu 1 Desember 2021 - 27 Juli 2022, hanya terdapat sebaran *raw data* milik msisdn00079 dan tidak ada *raw data* milik msisdn00080. Hal ini

menunjukkan bahwa selama rentang waktu tersebut, *subscribers* hanya menggunakan satu nomor/kartu sim yaitu msisdn00079. Selain itu, terlihat juga bahwa sebaran *raw data* berupa titik-titik records milik msisdn00079 terkonsentrasi di wilayah Sumatera Barat, Lampung, dan DKI Jakarta.

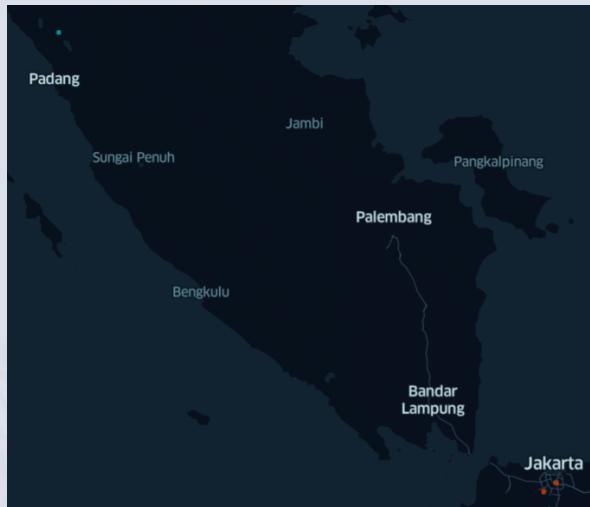
**Sebaran *Raw Data* msisdn00079 (Hijau) dan msisdn00080 (Merah) Selama
Rentang Waktu 1 Desember 2021 - 30 November 2022**



Gambar 4.43 Sebaran *Raw Data* msisdn00079 (Hijau) dan msisdn00080 (Merah) Rentang Waktu 1 Desember 2021 - 30 November 2022

Jika rentang waktu diperpanjang hingga 30 November 2022, maka mulai muncul sebaran *raw data* milik msisdn00080. Hal ini menunjukkan bahwa *subscribers* mulai menggunakan nomor/kartu sim msisdn00080 sejak tanggal 28 Juli 2022 yang berakibat pada berbedanya periode waktu penggunaan antara msisdn00079 dan msisdn00080. Terlihat juga bahwa sebaran *raw data* berupa titik-titik records milik msisdn00080 cenderung terkonsentrasi di wilayah DKI Jakarta, khususnya Kota Jakarta Timur. Selanjutnya, diimplementasikan algoritma AMDA untuk menentukan lokasi *home* masing-masing nomor sehingga diperoleh kandidat lokasi *home* seperti yang ditunjukkan pada visualisasi di bawah ini.

Hasil Penentuan Kandidat Lokasi *Home* msisdn00079 (Hijau) dan msisdn00080 (Merah) Menggunakan AMDA



Gambar 4.44 Hasil Penentuan Kandidat Lokasi *Home* msisdn00079 (Hijau) dan msisdn00080 (Merah) Menggunakan AMDA

Berdasarkan visualisasi di atas, terlihat bahwa msisdn00079 memiliki kandidat lokasi *home* di Provinsi Sumatera Barat, sedangkan msisdn00080 memiliki kandidat lokasi *home* di DKI Jakarta. Hasil akhir prediksi lokasi *home* dari kandidat lokasi *home* tersebut adalah sebagai berikut.

Tabel 4.15 Prediksi Lokasi *Home* msisdn00079 dan msisdn00080 dengan AMDA

msisdn	home_provinsi	home_kabupaten	provider
msisdn00079	SUMATERA BARAT	KABUPATEN TANAH DATAR	Telkomsel
msisdn00080	DKI JAKARTA	KOTA JAKARTA TIMUR	Indosat

Berdasarkan Tabel 4.15, dapat diketahui bahwa *subscribers* tersebut memiliki lokasi *home* yang berbeda untuk masing-masing nomor. msisdn00079 diprediksi memiliki lokasi *home* di Kabupaten Tanah Datar, Sumatera Barat, sedangkan msisdn00080 diprediksi memiliki lokasi *home* di Kota Jakarta Timur, DKI Jakarta.



Kedua kasus di atas dapat menyebabkan kesalahan dalam pengelompokan *subscribers* berdasarkan lokasi *home* yang sama. Seharusnya dua nomor yang digunakan dalam satu *handphone* utama setiap *subscribers* memiliki lokasi *home* yang sama dan dikelompokkan menjadi satu kelompok. Akan tetapi, kedua nomor tersebut dikelompokkan ke dalam kelompok yang berbeda karena lokasi *home* yang berbeda.

Penelitian ini hanya berfokus terhadap kinerja dari algoritma yang akan dibangun setelah tahapan pengelompokan *subscribers* dengan lokasi *home* yang sama, tanpa turut memperhitungkan keakuratan dari metode AMDA dalam memprediksi lokasi *home*. Oleh karena itu, hasil prediksi dan pengelompokan lokasi *home* yang sama diasumsikan sudah tepat tanpa ada salah pengelompokan karena lokasi *home subscribers* yang berbeda.

4.4.4 Periode Waktu Analisis

Setelah dilakukan pengelompokan *subscribers* dengan lokasi *home* yang sama, langkah selanjutnya adalah menentukan periode atau rentang waktu dalam menganalisis *subscribers* yang merupakan entitas *family grouping* dan *double counting*. Penelitian yang dilakukan oleh Ingkadijaya, dkk pada tahun 2015 menjelaskan bahwa rombongan keluarga semakin marak melakukan perjalanan wisata ke berbagai destinasi wisata, terutama pada *weekend*, *long weekend*, dan libur panjang. Hal tersebut juga senada dengan hasil penelitian yang dilakukan oleh Hehanussa, dkk pada tahun 2022 yang mengungkap bahwa kawasan wisata Saung Eling banyak dikunjungi oleh keluarga terutama pada hari libur dan akhir pekan (*weekend*).

Kedua penelitian tersebut menyimpulkan bahwa anggota keluarga memiliki kecenderungan untuk menghabiskan waktu di rumah atau bepergian ke tempat wisata secara bersama-sama ketika *weekend* atau

liburan. Hal tersebut lah yang menjadi dasar bagi penelitian ini untuk menggunakan periode waktu hari-hari *weekend* (Sabtu - Minggu) dan hari-hari libur lainnya dalam algoritma *family grouping*. Sementara itu, untuk algoritma *double counting*, periode waktu yang digunakan adalah hari-hari *weekdays* (Senin - Jumat) dengan asumsi bahwa aktivitas yang dilakukan pada hari kerja (*weekdays*) dapat memberikan pemisahan yang jelas dengan aktivitas yang dilakukan pada hari-hari *weekend*. Dengan begitu, diharapkan tidak ada tumpang tindih terkait periode waktu yang digunakan serta aktivitas seluler yang terkandung di dalamnya.

Untuk memperkuat alasan penggunaan waktu referensi tersebut, dilakukan penghitungan jarak antar *staypoint* antar nomor pada sampel terpilih menggunakan formula *haversine* pada periode waktu *weekdays* dan *weekend*. Sampel yang digunakan adalah lima *family* nomor *family grouping* dan lima pasangan nomor *double counting* yang terpilih secara acak. Hasil penghitungan dapat terlihat pada grafik di bawah ini.

Jarak Antar *Staypoint* pada Lima Sampel *Family Nomor Family Grouping*





Gambar 4.45 Jarak Antar *staypoint* pada Lima Sampel *Family* Nomor *Family Grouping*

Berdasarkan Gambar 4.45, Terlihat bahwa dari kelima sampel *family* nomor *family grouping*, seluruhnya memiliki rata-rata jarak antar *staypoint* yang lebih kecil ketika periode *weekends* dibandingkan dengan *weekdays*. Hal ini turut memperkuat alasan penggunaan periode waktu *weekends* untuk mengidentifikasi *subscribers* yang termasuk ke dalam satu keluarga (*family*) dalam algoritma *family grouping*.

Jarak Antar *Staypoint* pada Lima Sampel Pasangan Nomor *Double Counting*





Gambar 4.46 Jarak Antar *Staypoint* pada Lima Sampel Pasangan Nomor *Double Counting*

Berdasarkan gambar 4.46, terlihat bahwa dari kelima sampel pasangan nomor *double counting*, seluruhnya memiliki rata-rata jarak antar *staypoint* yang lebih kecil ketika periode *weekdays* dibandingkan dengan *weekends*. Hal ini turut memperkuat alasan penggunaan periode waktu *weekdays* untuk mengidentifikasi *subscribers* yang terhitung ganda (*double counting*) dalam algoritma *double counting*.

4.4.5 Eliminasi Pergerakan Tidak Wajar

Pergerakan tidak wajar adalah pergerakan yang memiliki kecepatan (*speed*) melebihi kecepatan yang dapat dilakukan manusia pada umumnya untuk berpindah antar tempat. Pada tahapan awal eksplorasi data MPD, pergerakan yang tidak wajar ini sering disebabkan oleh adanya *lag record*. *Lag record* adalah suatu kondisi dimana waktu pencatatan posisi perangkat seluler tidak sesuai dengan waktu sebenarnya di mana perangkat tersebut berada. Hal ini bisa terjadi karena beberapa alasan, seperti jaringan yang sibuk atau terbatasnya sumber daya BTS (*Base Transceiver Station*) yang digunakan untuk menerima sinyal perangkat seluler.

Lag Record dapat diilustrasikan sebagai berikut: misalkan terdapat sebuah perangkat seluler yang bergerak dari daerah A menuju daerah B pada pukul 9 pagi. Namun, karena ketidakakuratan pendektsian BTS atau adanya kendala jaringan, *record* MPD yang tercatat menunjukkan bahwa perangkat seluler tersebut masih berada di daerah A pada pukul 9 pagi, padahal seharusnya perangkat seluler itu sudah berada di daerah B pada

waktu tersebut. Dalam hal ini, dapat dikatakan terjadi *lag record* antara waktu sebenarnya (perangkat seluler berada di daerah B) dan waktu pencatatan posisinya (perangkat seluler masih tercatat di daerah A). *Lag record* dapat menyebabkan ketidakakuratan dalam analisis MPD, seperti pada estimasi kecepatan perjalanan, durasi perjalanan, atau pemodelan pola pergerakan pengguna. Oleh karena itu, penting untuk dilakukan eliminasi pergerakan (kecepatan) tidak wajar yang disebabkan oleh adanya *lag record*.

Proses eliminasi kecepatan yang tidak wajar dilakukan dengan menghitung kecepatan antara dua *record* secara berurutan, kemudian mengeliminasinya apabila terdapat kecepatan yang lebih dari 100 km/jam. Batasan kecepatan 100 km/jam ini menggunakan asumsi batas kecepatan maksimum berkendara pada jalan bebas hambatan menurut Peraturan Menteri Perhubungan Nomor PM 111 Tahun 2015.

Sebelumnya, pada tahapan AMDA juga terdapat proses eliminasi *fast movers* dengan menghapus *record* yang memiliki kecepatan di atas 10 km/jam. Penggunaan batasan 10 km/jam ini didasari pada asumsi rata-rata kecepatan manusia ketika berlari dan bertujuan agar memperoleh titik-titik koordinat yang benar-benar *stay* di suatu tempat agar mempermudah proses pembentukan *anchor* sebagai *usual environment* (Putra et al., 2021). Namun dalam analisis *family grouping* dan *double counting*, titik-titik koordinat yang dibutuhkan tidak hanya titik koordinat yang hanya *stay* di suatu tempat, melainkan juga titik-titik koordinat yang merepresentasikan adanya pergerakan/perjalanan dari suatu tempat ke tempat lainnya. Oleh karena itu, setelah nomor-nomor dikelompokkan berdasarkan kesamaan lokasi *home*, data MPD yang digunakan pada langkah-langkah selanjutnya adalah data MPD baru (bukan data MPD dari hasil *processing* AMDA). Kemudian dilakukan eliminasi pergerakan yang tidak wajar pada data MPD baru tersebut, yaitu pergerakan yang kecepatannya melebihi 100 km/jam. Berikut adalah contoh proses pengeliminasian pergerakan yang tidak wajar.

Tabel 4.16 Data MPD dengan Data Pergerakan yang Tidak Wajar

msisdn	datetime	source	latitude	longitude	Jarak haversine (km)	waktu (jam)	Kecepatan (km/jam)
msisdn00 xxx	12/10/2021 9:24	upcc	-6.16333	106.84			
msisdn00 xxx	12/10/2021 10:24	upcc	-6.16444	106.83583	0.477239	1.001389	0.4765
msisdn00 xxx	12/10/2021 11:24	upcc	-6.16444	106.83583	0	1.000556	0
msisdn00 xxx	12/10/2021 13:24	upcc	-6.20239	106.819	4.611796	2.000833	2.3049
msisdn00 xxx	12/10/2021 13:35	chg	-6.20239	106.819	0	0.184722	0
msisdn00 xxx	12/10/2021 14:15	chg	-6.26222	106.97889	18.88454	0.661111	28.564
msisdn00 xxx	12/10/2021 14:25	upcc	-6.16333	106.84	18.88473	0.172778	109.30

Baris (*record*) yang berwarna merah pada tabel di atas merupakan *record* yang akan dieliminasi karena dikategorikan sebagai pergerakan yang tidak wajar karena memiliki kecepatan lebih dari 100 km/jam.



4.4.6 Pembentukan *Staypoint*

Staypoint dibentuk dari kumpulan lokasi pelanggan yang berdekatan pada periode waktu tertentu yang level *raw data*-nya bersumber baik dari CDR dan/atau LBA/LBS yang kemudian dihitung sebagai titik tengah dari kumpulan titik lokasi tersebut (Dang et al., 2017). Menurut BPS (2021), *staypoint* adalah lokasi titik menetap pelanggan yang merupakan hasil penyederhanaan dan transformasi data transaksi dari lokasi BTS. Walaupun dilakukan penyederhanaan titik-titik lokasi, *staypoint* tetap merepresentasikan pergerakan pelanggan sesuai dengan kondisi pada level aslinya. Berdasarkan penjelasan tersebut, pada penelitian ini, *staypoint* mengacu pada lokasi yang merepresentasikan tempat menetap pelanggan (*subscribers*) pada rentang waktu tertentu dan dibentuk dari kumpulan lokasi pelanggan yang berdekatan pada periode waktu tersebut.

Algoritma pembentukan *staypoint* dimulai dengan membagi hari menjadi dua interval yang digunakan untuk membatasi perhitungan dengan membagi data yang tercatat pada waktu *work* dan waktu *home*. Waktu *work* yang digunakan adalah antara pukul 07.00 hingga 19.00, sementara waktu *home* adalah kebalikannya yaitu dari pukul 19.00 hingga 07.00. Pembagian interval ini mengikuti penelitian dari Putra et.al. (2021) yang menetapkan bahwa lokasi rumah adalah lokasi dengan rata-rata pencatatan waktu di bawah jam 7 dan lokasi kerja adalah lokasi dengan rata-rata pencatatan waktu di atas atau sama dengan jam 7.

Setelah data dibersihkan dan dilakukan eliminasi pergerakan yang tidak wajar, dilakukan penghitungan jumlah kemunculan tiap koordinat lokasi yang berbeda di setiap harinya untuk seluruh nomor dan untuk setiap interval. Koordinat lokasi yang berbeda dari nomor tersebut akan disebut sebagai *anchor*. Umumnya, akan ada beberapa *anchor* dalam satu hari, bahkan dalam satu interval. *Anchor* dengan frekuensi terbanyak dalam satu interval disebut sebagai kandidat utama dan secara otomatis

akan menjadi *staypoint*. Sementara itu, *anchor* lainnya akan memiliki dua kemungkinan, yakni dieliminasi ketika memiliki jarak kurang dari 500 meter dari kandidat utama atau dibiarkan dan menjadi *staypoint* baru (Putra et al., 2021) sehingga dalam satu interval bisa terdapat lebih dari satu *staypoint*. Tahapan pembentukan *staypoint* ini bertujuan untuk mengurangi waktu dan beban komputasi.

Sebagai contoh, akan ditentukan *staypoint* berdasarkan data pada Tabel 4.17. Pertama-tama, data perlu dibagi menjadi dua interval berdasarkan waktu pencatatan (*datetime*), yakni interval 1 (07.00-18.59) dan Interval 2 (19.00-06.59). Pada interval 1, terdapat dua titik koordinat (*anchor*), yaitu koordinat 1 (-6.24925, 106.90133) yang direpresentasikan dengan baris warna kuning dan koordinat 2 (-6.2501, 106.90658) yang direpresentasikan dengan baris warna biru. Koordinat 1 muncul sebanyak tiga kali dan koordinat 2 muncul sebanyak enam kali. Dengan begitu, koordinat 2 akan dijadikan sebagai kandidat utama.

Tabel 4.17 Contoh Pembentukan *Staypoint*

datetime	msisdn	source	latitude	longitude	provinsi	kabupaten	interval
11/02/2021 01:09	msisdn 00xxx	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 01:11	msisdn 00xxx	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	INTER-
11/02/2021 03:10	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	VAL 2
11/02/2021	msisdn	upcc	-6.24925	106.90133	DKI	Kota	



05:13		00xxx				Jakarta	Jakarta	
								Timur
11/02/2021 07:14	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 09:32	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 10:32	msisdn 00xxx	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 11:10	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 12:47	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	INTER-	
11/02/2021 13:12	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur		VAL 1
11/02/2021 15:10	msisdn 00xxx	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 17:01	msisdn 00xxx	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur		

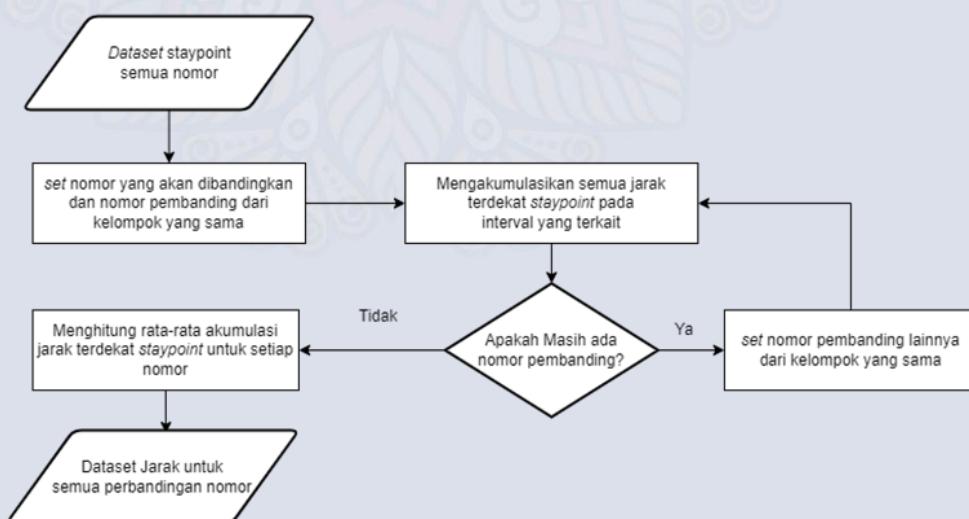
11/02/2021 17:13	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 19:12	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 20:42	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 21:03	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 23:07	msisdn 00xxx	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur

Setelah kandidat utama ditentukan, selanjutnya akan dihitung jarak *anchor* lainnya dengan kandidat utama. Jarak *Haversine* antara koordinat 1 (-6.24925, 106.90133) dan koordinat 2 (-6.2501, 106.90658) adalah 0.59 km. Karena jarak koordinat 2 (sebagai *anchor* lainnya) terhadap koordinat 1 (sebagai kandidat utama) lebih dari 0,5 km, maka koordinat 2 tidak akan dieliminasi dan akan dijadikan sebagai *staypoint*. Sehingga pada interval 1 terdapat dua *staypoint*, yakni koordinat 1 (-6.24925, 106.90133) dan koordinat 2 (-6.2501, 106.90658).

4.4.7 Pembangunan Algoritma dengan Kombinasi Jarak Terdekat antar Staypoint

Salah satu ukuran kesamaan (*similarity measure*) adalah dengan memanfaatkan penghitungan jarak (*distance*). Ide dasar pada algoritma ini adalah mengukur jarak antar *staypoint* pada setiap nomor *subscribers* yang berada di kelompok yang sama (memiliki lokasi *home* yang sama) menggunakan jarak *Haversine*. Jarak *Haversine* dipilih karena memperhitungkan bahwa bumi bukanlah sebuah bidang datar namun adalah sebuah bidang yang memiliki derajat kelengkungan (Putra et al., 2021). Berdasarkan penghitungan jarak *Haversine* tersebut, diharapkan akan didapatkan pola hubungan antar *staypoint* pada nomor-nomor entitas *family grouping* dan *double counting*. Jika beberapa nomor memiliki jarak yang relatif dekat, maka hal tersebut dapat mengindikasikan bahwa nomor-nomor tersebut memiliki hubungan satu sama lain seperti hubungan keluarga (*family grouping*) dan bahkan dapat merepresentasikan *subscribers* yang sama (*double counting*). Langkah-langkah yang dilakukan pada tahap ini terlihat pada gambar 4.47.

Flowchart Tahapan Perhitungan Jarak antar Staypoint dengan Haversine



Gambar 4.47 Flowchart Perhitungan Jarak antar Staypoint dengan Haversine

Setelah mendapatkan *staypoint* pada masing-masing interval di setiap harinya untuk seluruh nomor dengan menerapkan langkah-langkah pembentukan *staypoint* pada subbab 4.4.6, langkah selanjutnya adalah membandingkan nomor-nomor pada kelompok sama. Contohnya, membandingkan msisdn00xxx1 dengan nomor-nomor lainnya di dalam kelompok yang sama. Proses perbandingan dilakukan dengan menghitung jarak *Haversine* antara *staypoint* milik msisdn001 dengan *staypoint* milik nomor lainnya (misal msisdn00xxx2) di interval dan hari yang bersesuaian. Berikut adalah contoh perhitungannya

Tabel 4.18 MSISDN 00xxx1 dengan 2 *Staypoint* pada Interval 1

datetime	msisdn	source	latitude	longitude	provinsi	kabupaten	interval
11/02/2021 01:09	msisdn 00xxx1	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 01:11	msisdn 00xxx1	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 03:10	msisdn 00xxx1	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	INTER- VAL 2
11/02/2021 05:13	msisdn 00xxx1	upcc	-6.24925	106.90133	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 07:14	msisdn 00xxx1	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	



09:32	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.24925	106.90133	DKI	Kota	
10:32	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	
11:10	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	INTER-
12:47	00xxx1				Jakarta	Jakarta	VAL 1
							Timur
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	
13:12	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.24925	106.90133	DKI	Kota	
15:10	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.24925	106.90133	DKI	Kota	
17:01	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	
17:13	00xxx1				Jakarta	Jakarta	
							Timur
11/02/2021	msisdn	upcc	-6.25010	106.90658	DKI	Kota	
19:12	00xxx1				Jakarta	Jakarta	
							Timur

11/02/2021 20:42	msisdn 00xxx1	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	INTER- VAL 2
11/02/2021 21:03	msisdn 00xxx1	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 23:07	msisdn 00xxx1	upcc	-6.25010	106.90658	DKI Jakarta	Kota Jakarta Timur	

Tabel 4.19 MSISDN 00xxx2 dengan 3 Staypoint pada Interval 1

datetime	msisdn	source	latitude	longitude	provinsi	kabupaten	interval
11/02/2021 01:12	msisdn 00xxx2	upcc	-6.55956	107.89661	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 03:10	msisdn 00xxx2	upcc	-6.55956	107.89661	DKI Jakarta	Kota Jakarta Timur	
11/02/2021 05:01	msisdn 00xxx2	upcc	-6.55956	107.89661	DKI Jakarta	Kota Jakarta Timur	INTER- VAL 2
11/02/2021 05:13	msisdn 00xxx2	upcc	-6.55956	107.89661	DKI Jakarta	Kota Jakarta Timur	
11/02/2021	msisdn	upcc	-6.55956	107.89661	DKI	Kota Jakarta	



07:12	00xxx2			Jakarta	Timur
11/02/2021 08:42	msisdn 00xxx2	upcc	-6.55956 107.89661	DKI Jakarta	Kota Jakarta Timur
11/02/2021 09:03	msisdn 00xxx2	upcc	-6.55956 107.89661	DKI Jakarta	Kota Jakarta Timur
11/02/2021 11:07	msisdn 00xxx2	upcc	-6.55956 107.89661	DKI Jakarta	Kota Jakarta Timur
11/02/2021 14:57	msisdn 00xxx2	upcc	-6.25010 106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 15:19	msisdn 00xxx2	upcc	-6.25010 106.90658	DKI Jakarta	Kota Jakarta Timur
11/02/2021 17:16	msisdn 00xxx2	upcc	-6.24925 106.90133	DKI Jakarta	Kota Jakarta Timur
11/02/2021 18:31	msisdn 00xxx2	upcc	-6.24925 106.90133	DKI Jakarta	Kota Jakarta Timur
11/02/2021 19:33	msisdn 00xxx2	upcc	-6.64335 108.05679	DKI Jakarta	Kota Jakarta Timur
11/02/2021	msisdn	upcc	-6.64335 108.05679	DKI	Kota

20:37	00xxx2					Jakarta	Jakarta	
							Timur	
11/02/2021 21:27	msisdn 00xxx2	upcc	-6.64335	108.05679	DKI Jakarta	Kota Jakarta Timur		INTER- VAL 2
11/02/2021 22:52	msisdn 00xxx2	upcc	-6.64335	108.05679	DKI Jakarta	Kota Jakarta Timur		
11/02/2021 23:50	msisdn 00xxx2	upcc	-6.64335	108.05679	DKI Jakarta	Kota Jakarta Timur		

Pada contoh diatas, terlihat bahwa msisdn00xxx1 memiliki total dua *staypoint* pada interval 1 dan msisdn00xxx2 memiliki total tiga *staypoint* pada interval 1, dengan rincian sebagai berikut:

1. SP1Nomor1 : (-6.25010,106.90658)
2. SP2Nomor1 : (-6.24925,106.90133)
3. SP1Nomor2 : (-6.55956,107.89661)
4. SP2Nomor2 : (-6.25010,106.90658)
5. SP3Nomor2 : (-6.24925,106.90133)

Langkah selanjutnya adalah menghitung jarak *Haversine* untuk masing-masing kombinasi *staypoint* diantara kedua nomor tersebut. Beberapa kombinasi *staypoint* yang mungkin diantaranya adalah:

1. $d_1 = \text{hav}(\text{SP1Nomor1}, \text{SP1Nomor2}) = 114811,6 \text{ m}$
2. $d_2 = \text{hav}(\text{SP1Nomor1}, \text{SP2Nomor2}) = 0 \text{ m}$
3. $d_3 = \text{hav}(\text{SP1Nomor1}, \text{SP3Nomor2}) = 588,6092 \text{ m}$
4. $d_4 = \text{hav}(\text{SP2Nomor1}, \text{SP1Nomor2}) = 115394,2 \text{ m}$
5. $d_5 = \text{hav}(\text{SP2Nomor1}, \text{SP2Nomor2}) = 588,6092 \text{ m}$



$$6. \ d6 = \text{hav}(\text{SP2Nomor1}, \text{SP3Nomor2}) = 0 \text{ m}$$

Berdasarkan hasil penghitungan jarak antara dua *staypoint* pada enam kombinasi yang ada, selanjutnya dipilih kombinasi yang memiliki jarak terdekat antara dua *staypoint*. Pada contoh perhitungan di atas, jarak terdekat adalah d_2 dan d_6 dengan jarak antar dua *staypoint* 0 m. Oleh karena itu, untuk Interval 1 pada msisdn00xxx1 dan msisdn00xxx2 di hari yang bersesuaian memiliki sekumpulan *staypoint* yang berjarak 0 m. Langkah penghitungan ini diterapkan untuk setiap interval di setiap harinya selama setahun untuk mendapatkan jarak *Haversine* terdekat dari sekumpulan *staypoint* pada interval-interval yang bersesuaian dari msisdn00xxx1 dan msisdn00xxx2. Selanjutnya, dilakukan penghitungan selisih jarak hasil dari penghitungan jarak haversine terdekat antar *staypoint* pada suatu interval dengan jarak haversine terdekat pada interval berikutnya. Misalnya, jarak haversine terdekat Interval 1, d_1 , adalah 0 meter dan jarak haversine terdekat Interval 2, d_2 , adalah 5 meter. Maka, Selisih jarak Interval 1 dan Interval 2 adalah 5 meter.

langkah tersebut dilakukan untuk interval selanjutnya secara berurutan. Kemudian jarak terdekat antar interval tersebut dirata-ratakan selama setahun. Perhitungan ini bertujuan untuk membentuk variabel baru yang menunjukkan rata-rata kestabilan jarak terdekat antar interval selama satu tahun pada nomor-nomor yang berada di kelompok yang sama.

Tabel 4.20 Contoh Rata-Rata Kestabilan Jarak Terdekat antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Family	Average_D
msisdn00xxx92	msisdn00xxx109	TRUE	0.015145
msisdn00xxx42	msisdn00xxx79	FALSE	9.926563
msisdn00xxx79	msisdn00xxx101	FALSE	8.224279

mslsdn00xxx42	mslsdn00xxx101	TRUE	0.000000
msisdn00xxx51	msisdn00xxx100	TRUE	0.140389
msisdn00xxx35	msisdn00xxx98	TRUE	0.140924

Setelah didapatkan variabel rata-rata kestabilan jarak terdekat antar nomor, untuk mengidentifikasi entitas yang termasuk ke dalam *family grouping* ataupun *double counting*, dilakukan pembangunan *threshold* rata-rata jarak kestabilan antar nomor yang dapat mengklasifikasikan mana entitas yang termasuk dan mana yang tidak termasuk dalam *family grouping* atau *double counting* dengan baik. Langkah-langkah pembangunan *threshold* ini akan dijelaskan pada subbab 4.4.10.

4.4.8 Pembangunan Algoritma *Hexagonal Hierarchical Geospatial Indexing System* (H3)

Hexagonal Hierarchical Geospatial Indexing System (H3) merupakan sistem pengindeksan geospasial untuk mengelompokkan data spasial berdasarkan *grid* yang berbentuk heksagonal atau segi enam. Gagasan utama metode ini adalah bahwa heksagon merupakan bentuk yang paling bermanfaat untuk analisis spasial dan dapat digunakan untuk membuat perhitungan yang konsisten (Jordan Bean, 2022). Hal tersebut disebabkan oleh bentuk heksagonal dapat membuat semua tetangga memiliki jarak yang sama (Uber Technologies, Inc., 2023).

Setiap heksagon pada H3 memiliki nomor identitas unik, yang disebut sebagai *H3 index*. *H3 index* dapat digunakan untuk memetakan lokasi geospasial pada permukaan bumi ke dalam sebuah heksagon dan dapat digunakan untuk menghitung jarak antara dua lokasi yang diwakili oleh dua heksagon yang berbeda. Sistem H3 memiliki struktur hierarkis berupa 15 resolusi yang memungkinkan identifikasi unik dan hubungan antara sel-sel heksagonal pada berbagai tingkat resolusi tersebut. Adanya



tingkat resolusi tersebut memungkinkan indeks dan pengambilan data geospasial yang efisien serta operasi spasial seperti agregasi dan filtrasi. Dengan menggunakan H3, pengguna dapat dengan mudah mengkonversi koordinat lintang dan bujur menjadi indeks H3, menemukan sel-sel tetangga, dan melakukan perhitungan spasial berdasarkan struktur *grid* heksagonal (Uber Technologies, Inc., 2023).

Berdasarkan keunggulan-keunggulan tersebut, maka sistem pengindeksan H3 dipilih sebagai pendekatan yang digunakan dalam membangun algoritma *family grouping* dan *double counting*. Selain itu, pembangunan algoritma *family grouping* dan *double counting* dengan pendekatan H3 ini dilakukan dengan menggunakan 2 metode, yaitu tanpa pembentukan *staypoint* dan dengan pembentukan *staypoint*.

4.4.3.1 Tanpa Pembentukan *Staypoint*

Setelah terlebih dahulu mengeliminasi pergerakan yang tidak wajar sesuai dengan tahapan yang sudah dijelaskan sebelumnya, langkah selanjutnya yang dilakukan adalah melakukan konversi atau pemetaan untuk setiap nomor ke dalam bentuk heksagon. Lokasi koordinat untuk masing-masing nomor dikonversi menggunakan sistem pengindeksan H3 menjadi heksagon dengan *cell id* yang unik. Berikut merupakan contoh hasil konversi koordinat lokasi menjadi indeks H3.

Tabel 4.21 Contoh Konversi Koordinat Lokasi Menjadi Indeks H3 (Resolusi 8)

Koordinat Lokasi (Lat, Long)	Indeks H3
(-6.25010, 106.90658)	888c107b19fffff
(-6.24925, 106.90133)	888c107b1bfffff
(-6.55956, 107.89661)	888c16b927fffff

(-6.64335, 108.05679)	888c168e69fffff
-----------------------	-----------------

Resolusi H3 yang disarankan berdasarkan hasil penelitian dari Kent Marten, Michael Johns, dan Menelaos Karavellas yang berjudul *Spatial Analytics at Any Scale With H3 and Photon* adalah resolusi 8 sampai dengan resolusi 12. Penelitian tersebut melakukan evaluasi terhadap berbagai resolusi yang dapat digunakan untuk pendekatan *hybrid H3-centric* dan juga *H3-Geometry* dengan mempertimbangkan resolusi 8, 10, dan 12. Namun, untuk mengetahui resolusi H3 terbaik yang dapat digunakan dalam membangun algoritma *family grouping* dan *double counting*, penelitian ini akan tetap menggunakan resolusi 1 sampai 12. Dari seluruh resolusi H3 tersebut, akan dipilih satu resolusi H3 terbaik dengan membandingkan performa antar resolusi.

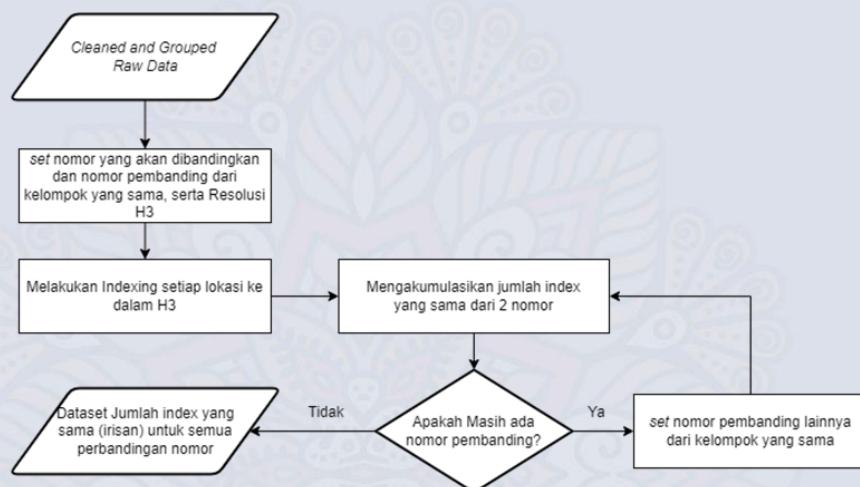
Misalkan pada langkah ini akan digunakan nomor dengan id msisdn00xxx1 dengan resolusi 8. Sebelum membandingkan heksagon antar nomor, akan dilakukan eliminasi indeks H3 yang memiliki duplikasi selama periode waktu analisis (Desember 2021 - November 2022) sehingga indeks H3 yang akan dibandingkan hanya indeks H3 yang unik.

Tabel 4.22 Contoh Indeks H3 yang Unik

Koordinat Lokasi (Lat, Long)	Indeks H3
(-6.25010, 106.90658)	888c107b19fffff
(-6.24925, 106.90133)	888c107b1bfffff
(-6.55956, 107.89661)	888c16b927fffff
(-6.64335, 108.05679)	888c168e69fffff
(-7.79510, 110.35873)	888d8cb819fffff

Setelah diperoleh indeks H3 yang unik, langkah selanjutnya adalah menghitung jumlah heksagon milik msisdn00xxx1 yang beririsan dengan heksagon milik nomor lainnya yang masih berada dalam kelompok yang sama. Tahapan yang dilakukan dalam melakukan penghitungan jumlah heksagon yang beririsan dapat terlihat pada gambar 4.48.

Flowchart Tahapan Penghitungan Jumlah Irisan H3 Tanpa Staypoint



Gambar 4.48 Flowchart Penghitungan Jumlah Irisan H3 Tanpa Staypoint

Misalkan msisdn00xxx1 memiliki sebaran heksagon sebagai berikut:

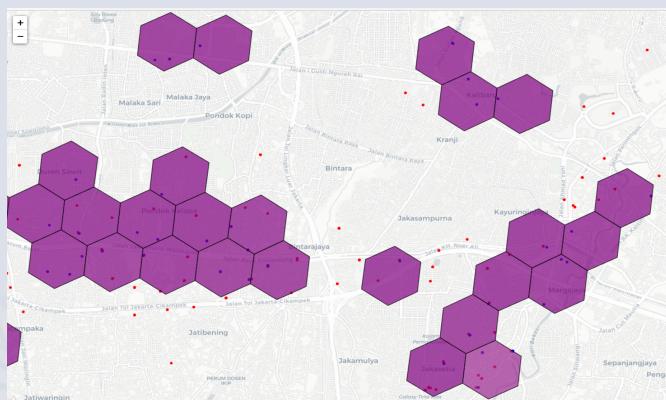
Visualisasi Sebaran Heksagon msisdn00xxx1



Gambar 4.49 Sebaran Heksagon msisdn00xxx1

Sementara itu, nomor lainnya (misal msisdn00xxx2) memiliki sebaran heksagon sebagai berikut:

Visualisasi Sebaran Heksagon msisdn00xxx2



Gambar 4.50 Sebaran Heksagon msisdn00xxx2

Jika kita meng-*overlay*-kan *layer* heksagon msisdn00xxx1 dengan *layer* heksagon msisdn00xxx2, akan didapatkan sebaran hexagon sebagai berikut:

Visualisasi Sebaran Heksagon msisdn00xxx1 dan msisdn00xxx2



Gambar 4.51 Sebaran Heksagon msisdn00xxx1 dan msisdn00xxx2

Dengan begitu, pada gambar 4.52, terlihat bahwa antara msisdn00xxx1 dengan msisdn00xxx2 memiliki jumlah heksagon yang beririsan sebanyak 19 heksagon.

Visualisasi Jumlah Heksagon Beririsan**Gambar 4.52** Jumlah Heksagon yang Beririsan

Langkah di atas akan dilakukan secara berulang dengan menghitung jumlah irisan heksagon antar nomor pada kelompok yang sama sampai mencakup semua kemungkinan kombinasi dua nomor pada kelompok tersebut. Setelah tahapan perhitungan jumlah irisan heksagon selesai, akan didapatkan sebuah variabel jumlah irisan heksagon untuk setiap dua nomor yang dibandingkan dalam kelompok yang sama. Contoh variabel tersebut dapat terlihat pada tabel di bawah ini.

Tabel 4.23 Jumlah Heksagon yang Beririsan antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Family	Jumlah_Irisan
msisdn00xxx92	msisdn00xxx109	TRUE	88
msisdn00xxx42	msisdn00xxx101	TRUE	201
msisdn00xxx42	msisdn00xxx79	FALSE	31
msisdn00xxx79	msisdn00xxx101	FALSE	15
mslsdn00xxx51	msisdn00xxx100	TRUE	70
mslsdn00xxx35	msisdn00xxx98	TRUE	104
msisdn00xxx18	msisdn00xxx21	TRUE	84

msisdn00xxx91	msisdn00xxx111	TRUE	94
msisdn00xxx91	msisdn00xxx110	TRUE	93
msisdn00xxx49	msisdn00xxx91	TRUE	67

Untuk menentukan entitas yang termasuk ke dalam *family grouping* atau *double counting*, akan ditentukan *threshold* berupa jumlah irisan heksagon yang dapat mengklasifikasikan mana entitas *double counting* atau *family grouping* dan mana yang bukan entitas *double counting* atau *family grouping* dengan baik. Langkah-langkah pembangunan *threshold* ini akan dijelaskan pada subbab 4.4.10.

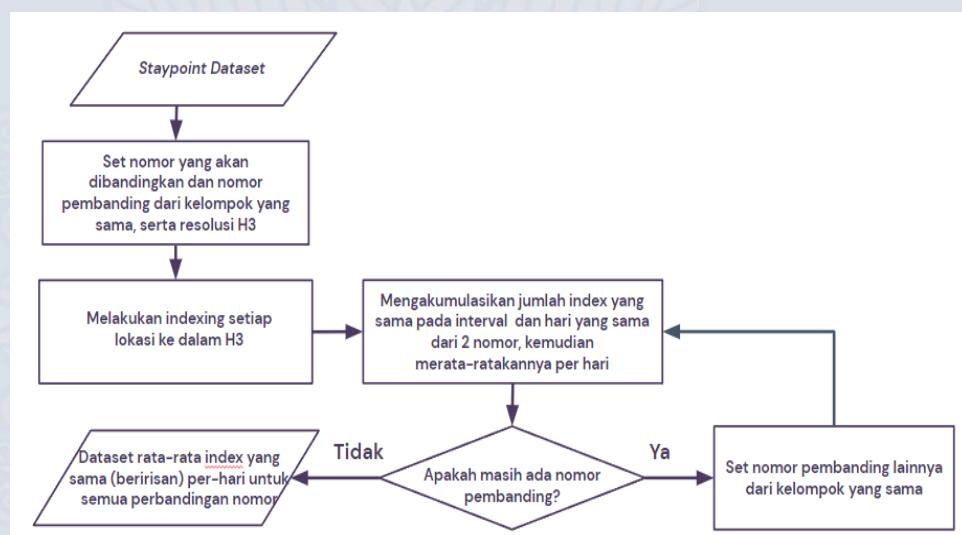
4.4.3.2 Dengan Pembentukan *Staypoint*

Perbedaan utama antara pembangunan algoritma H3 tanpa pembentukan *staypoint* dan dengan pembentukan *staypoint* adalah terkait koordinat lokasi yang akan dikonversi ke dalam bentuk indeks H3. Pada algoritma H3 tanpa pembentukan *staypoint*, koordinat lokasi yang dikonversi adalah semua koordinat lokasi yang dimiliki oleh suatu nomor. Sedangkan pada algoritma H3 dengan pembentukan *staypoint*, koordinat lokasi yang dikonversi merupakan *staypoint* yang telah dibentuk dengan langkah-langkah yang dijelaskan pada subbab 4.4.6. Perbedaan lainnya adalah terkait perbandingan indeks H3 yang beririsan antar dua nomor. Pada algoritma H3 tanpa pembentukan *staypoint*, indeks H3 yang dibandingkan adalah seluruh indeks H3 selama periode waktu analisis (Desember 2021 - November 2022). Sementara itu, pada algoritma H3 dengan pembentukan *staypoint*, indeks H3 yang dibandingkan hanya pada interval waktu yang bersesuaian selama periode waktu analisis, dengan catatan akan dilakukan eliminasi terlebih dahulu terhadap indeks H3 yang terdapat duplikasi pada setiap interval.

Setelah dilakukan perbandingan dan perhitungan jumlah indeks H3 yang beririsan, jumlah indeks H3 yang beririsan tersebut akan

dirata-ratakan per harinya. Penghitungan rata-rata ini hanya memperhitungkan hari yang terdapat minimal satu indeks H3 yang beririsan antar dua nomor agar terdapat standar perbandingan yang sama antara setiap pasangan nomor yang dibandingkan tanpa dipengaruhi oleh jumlah *records* yang dimiliki nomor tersebut. Tahapan yang dilakukan untuk menghitung jumlah heksagon yang beririsan dengan pembentukan *staypoint* terlihat pada gambar 4.53.

Flowchart Tahapan Penghitungan Rata-Rata Jumlah Irisan H3 dengan Staypoint



Gambar 4.53 Flowchart Tahapan Penghitungan Jumlah Irisan H3 dengan Staypoint

Berikut ini dicontohkan perbedaan penghitungan jumlah irisan indeks H3 unik antar dua nomor tanpa pembentukan *staypoint* dan dengan pembentukan *staypoint*.

Tabel 4.24 Contoh Penghitungan Irisan Indeks H3 Tanpa Pembentukan *Staypoint*
msisdn00xxx1

datetime	Koordinat Lokasi	Indeks H3	Indeks H3 Unik
12/10/2021 1:16	(-6.25010, 106.90658)	888c107b19fffff	888c107b19fffff
12/10/2021 3:21	(-6.24925, 106.90133)	888c107b1bfffff	
12/10/2021 4:24	(-6.24925, 106.90133)	888c107b1bfffff	888c107b1bfffff
12/10/2021 6:24	(-6.55956, 107.89661)	888c16b927fffff	
12/10/2021 8:24	(-6.55956, 107.89661)	888c16b927fffff	
12/10/2021 10:24	(-6.55956, 107.89661)	888c16b927fffff	888c16b927fffff
12/10/2021 12:15	(-6.64335, 108.05679)	888c168e69fffff	888c168e69fffff
12/10/2021 13:25	(-7.79510, 110.35873)	888d8cb819fffff	
12/10/2021 16:25	(-7.79510, 110.35873)	888d8cb819fffff	888d8cb819fffff
12/10/2021 17:26	(-7.79510, 110.35873)	888d8cb819fffff	



Tabel 4.25 Contoh Penghitungan Irisan Indeks H3 Tanpa Pembentukan *Staypoint*
msisdn00xxx2

datetime	Koordinat Lokasi	Indeks H3	Indeks H3 Unik
12/10/2021 1:56	(-6.24925, 106.90133)	888c107b1bfffff	
12/10/2021 2:21	(-6.24925, 106.90133)	888c107b1bfffff	888c107b1bfffff
12/10/2021 3:44	(-6.24925, 106.90133)	888c107b1bfffff	
12/10/2021 6:10	(-6.55956, 107.89661)	888c16b927fffff	888c16b927fffff
12/10/2021 9:13	(-6.55956, 107.89661)	888c168e69fffff	
12/10/2021 11:58	(-6.64335, 108.05679)	888c168e69fffff	888c168e69fffff
12/10/2021 14:15	(-6.64335, 108.05679)	888c168e69fffff	
12/10/2021 16:35	(-6.64335, 108.05679)	888c168e69fffff	
12/10/2021 17:25	(-7.79510, 110.35873)	888d8cb819fffff	888d8cb819fffff
12/10/2021 18:46	(-7.79510, 110.35873)	888d8cb819fffff	

Berdasarkan tabel 4.24, diketahui bahwa msisdn00xxx1 pada tanggal 12/10/2021 memiliki jumlah indeks H3 unik sebanyak 5 indeks. Sedangkan berdasarkan tabel 4.25, diketahui bahwa msisdn00xxx2 pada tanggal 12/10/2021 memiliki jumlah indeks H3 unik sebanyak 4 indeks. Jika msisdn00xxx1 dan msisdn00xxx2 dibandingkan, jumlah indeks H3 yang beririsan adalah sebanyak 4 indeks, yaitu indeks 888c107b1bfffff, 888c16b927fffff, 888c168e69fffff dan 888d8cb819fffff.

Tabel 4.26 Contoh Penghitungan Irisan Indeks H3 dengan Pembentukan
Staypoint msisdn00xxx1

datetime	Interval	Koordinat Lokasi	Indeks H3	Indeks H3 Unik
12/10/2021 1:16		(-6.25010, 106.90658)	888c107b19fffff	888c107b19fffff
12/10/2021 3:21	Interval 1	(-6.24925, 106.90133)	888c107b1bfffff	888c107b1bfffff
12/10/2021 4:24		(-6.24925, 106.90133)	888c107b1bfffff	
12/10/2021 6:24		(-6.55956, 107.89661)	888c16b927fffff	888c16b927fffff
12/10/2021 8:24	Interval 2	(-6.55956, 107.89661)	888c16b927fffff	888c16b927fffff
12/10/2021 10:24		(-6.55956, 107.89661)	888c16b927fffff	



12/10/2021 12:15	(-6.64335, 108.05679)	888c168e69fffff	888c168e69fffff
12/10/2021 13:25	(-7.79510, 110.35873)	888d8cb819fffff	
12/10/2021 16:25	(-7.79510, 110.35873)	888d8cb819fffff	88d8cb819fffff
12/10/2021 17:26	(-7.79510, 110.35873)	888d8cb819fffff	

Tabel 4.27 Contoh Penghitungan Irisan Indeks H3 dengan Pembentukan
Staypoint msisdn00xxx2

datetime	Interval	Koordinat Lokasi	Indeks H3	Indeks H3 Unik
12/10/2021 1:56		(-6.24925, 106.90133)	888c107b1bfffff	
12/10/2021 2:21	Interval 1	(-6.24925, 106.90133)	888c107b1bfffff	888c107b1bfffff
12/10/2021 3:44		(-6.24925, 106.90133)	888c107b1bfffff	

12/10/2021 6:10	(-6.55956, 107.89661)	888c16b927fffff	888c16b927fffff
12/10/2021 9:13	(-6.55956, 107.89661)	888c168e69fffff	888c168e69fffff
12/10/2021 11:58	(-6.64335, 108.05679)	888c168e69fffff	
12/10/2021 14:15	(-6.64335, 108.05679)	888c168e69fffff	888c168e69fffff
12/10/2021 16:35	(-6.64335, 108.05679)	888c168e69fffff	
12/10/2021 17:25	(-7.79510, 110.35873)	888d8cb819fffff	888d8cb819fffff
12/10/2021 18:46	(-7.79510, 110.35873)	888d8cb819fffff	

Berdasarkan tabel 4.26, diketahui bahwa msisdn00xxx1 pada tanggal 12/10/2021 memiliki jumlah indeks H3 unik sebanyak 3 indeks pada interval 1 dan 3 indeks pada interval 2. Sedangkan berdasarkan tabel 4.27, msisdn00xxx2 pada tanggal 12/10/2021 memiliki jumlah indeks H3 unik sebanyak 2 indeks pada interval 1 dan 2 indeks pada interval 2. Jika msisdn00xxx1 dan msisdn00xxx2 dibandingkan, jumlah indeks H3 yang beririsan adalah sebanyak 4 indeks yang terdiri dari 2 indeks beririsan pada interval 1 (indeks 888c107b1bfffff dan 888c16b927fffff) serta 2



indeks beririsan pada interval 2 (indeks 888c168e69fffff dan 888d8cb819fffff). Jumlah indeks H3 yang beririsan tersebut kemudian dirata-ratakan sesuai dengan jumlah hari yang terdapat minimal 1 indeks H3 yang beririsan sehingga akan didapatkan sebuah variabel berupa rata-rata jumlah irisan heksagon per hari untuk setiap dua nomor yang dibandingkan dalam kelompok yang sama.

Tabel 4.28 Rata-rata Jumlah Heksagon yang Beririsan per Hari antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Family	Rata2_Jumlah_Irisan
msisdn00xxx92	msisdn00xxx109	TRUE	2,23
msisdn00xxx42	msisdn00xxx101	TRUE	2,01
msisdn00xxx42	msisdn00xxx79	FALSE	1,07
msisdn00xxx79	msisdn00xxx101	FALSE	1,14
mslsdn00xxx51	msisdn00xxx100	TRUE	2,36
mslsdn00xxx35	msisdn00xxx98	TRUE	2,11
msisdn00xxx18	msisdn00xxx21	TRUE	2,39
msisdn00xxx91	msisdn00xxx111	TRUE	1,98
msisdn00xxx91	msisdn00xxx110	TRUE	2,15
msisdn00xxx49	msisdn00xxx91	TRUE	2,27

Untuk menentukan entitas yang termasuk ke dalam *family grouping* atau *double counting*, akan ditentukan *threshold* (ambang batas) berupa rata-rata jumlah irisan heksagon per hari yang dapat mengklasifikasikan mana entitas *double counting* atau *family grouping* dan mana yang bukan entitas *double counting* atau *family grouping* dengan

baik. Langkah-langkah pembangunan *threshold* ini akan dijelaskan pada subbab 4.4.10.

4.4.9 Pembangunan Algoritma dengan *Dynamic Time Warping* dan *Frechet Distance*

Sulitnya mencari penelitian terkait pengukuran similaritas pada data pasif MPD membuat penelitian ini mencoba menggunakan metode pengukuran similaritas yang diterapkan pada MPD *active*. Beberapa metode pengukuran yang dapat digunakan untuk mengukur similaritas lintasan (trajectories) MPD *active* adalah *Longest Common Subsequence* (LCSS), *Frechet Distance*, *Dynamic Time Warping* (DTW), dan *Edit Distance* (Toohey K. & Matt D., 2015). Berdasarkan penelitian yang dilakukan oleh Kevin Toohey & Matt Duckham (2015) yang berjudul “Trajectory Similarity Measures”, terdapat korelasi positif yang sangat kuat (*correlation coefficients*: 0.9587) antara DTW dan *Frechet Distance* serta korelasi negatif yang cukup kuat (*correlation coefficients*: -0.8340) antara LCSS dan *Edit Distance*. Perbedaan korelasi yang sedikit lebih rendah pada LCSS dan *Edit Distance* jika dibandingkan dengan korelasi DTW dan *Frechet Distance* kemungkinan besar disebabkan oleh variasi yang cukup besar pada panjang lintasan. Hal tersebut disebabkan oleh DTW dan *Frechet Distance* yang tidak dipengaruhi oleh panjang lintasan sehingga varians yang besar pada panjang lintasan tidak mempengaruhi hasil penghitungan similaritas antar lintasan (Toohey K. & Matt D., 2015). Berdasarkan penelitian tersebut, pada penelitian ini akan digunakan DTW dan *Frechet Distance* sebagai metode pengukuran similaritas antar lintasan sebagai pendekatan untuk membangun algoritma *family grouping* dan *double counting*.

DTW adalah ukuran kesamaan (similarity measures) yang mengandalkan titik pencocokan (matching points) dalam lintasan. DTW mengukur jarak antara dua lintasan dengan panjang masing-masing m dan k serta memiliki kompleksitas $O(mk)$ (Toohey K. & Matt D., 2015).



Dengan menggunakan DTW, lintasan "dibengkokkan" secara non-linier untuk mengukur kesamaan antar lintasan sambil memungkinkan pengambilan sampel (titik koordinat yang akan dibandingkan) yang bervariasi (Yuan, 2008). Dengan DTW, satu titik pada suatu lintasan dapat dicocokkan dengan beberapa titik pada lintasan lainnya sehingga DTW dapat bekerja dengan baik pada lintasan dengan panjang yang berbeda dan bahkan dengan pengambilan sampel yang sangat bervariasi (Toohey K. & Matt D., 2015).

Namun, *outlier* dapat sangat mempengaruhi hasil dari metode DTW karena setiap titik koordinat dari kedua lintasan harus memiliki setidaknya satu pasang kecocokan (Toohey K. & Matt D., 2015). Untuk mengatasi permasalahan tersebut, sebelum mengimplementasikan DTW, dilakukan pembentukan *staypoint* terlebih dahulu untuk setiap lintasan dengan langkah-langkah yang telah dijelaskan sebelumnya. Selain itu, jarak haversine digunakan sebagai *distance function* pada penelitian ini karena memperhitungkan bahwa bumi bukanlah sebuah bidang datar namun adalah sebuah bidang yang memiliki derajat kelengkungan (Putra, 2019).

Sama halnya dengan DTW, *Frechet Distance* juga dapat diterapkan untuk mengukur tingkat kesamaan antara dua kurva kontinu berarah dan juga dua lintasan diskrit. Metode ini memungkinkan perhitungan jarak antara dua lintasan dengan panjang m dan k , dengan kompleksitas kasus terburuknya adalah $O((m^2k+k^2m) \log mk)$ (Toohey K. & Matt D., 2015). Pada *Frechet Distance*, titik-titik koordinat lintasan tidak dicocokkan secara bersamaan sehingga metode ini dapat bekerja dengan baik bahkan dengan tingkat pengambilan sampel dan panjang lintasan yang sangat bervariasi (Toohey K. & Matt D., 2015). Namun, hasil dari *Frechet Distance* dapat sangat dipengaruhi oleh *outlier* jika *outlier* tidak dihapus sebelum melakukan perhitungan. Oleh karena itu, sebelum mengimplementasikan *Frechet Distance*, pembentukan *staypoint* juga

dilakukan terlebih dahulu untuk setiap lintasan. Adapun jarak haversine, juga digunakan sebagai *distance function* dalam Frechet Distance.

Flowchart Tahapan Metode DTW dan Frechet Distance



Gambar 4.54 Flowchart Tahapan Metode DTW dan Frechet Distance

Tahapan awal dalam algoritma ini adalah melakukan pembentukan *staypoint* untuk setiap nomor dengan langkah-langkah seperti yang sudah dijelaskan pada subbab 4.4.6. Setelah dilakukan pembentukan *staypoint*, tahapan selanjutnya adalah melakukan ekstraksi lintasan spasial untuk setiap nomor dalam rentang waktu periode analisis (Desember 2021-November 2022). Lintasan spasial terdiri dari sekumpulan list *staypoint* yang terurut berdasarkan datetime. Lintasan spasial dimulai dari *staypoint* pertama yang menunjukkan lokasi *home* dari nomor tersebut dan diakhiri dengan staypoint terakhir yang menunjukkan lokasi *home* yang sama. Berikut adalah contoh proses ekstraksi lintasan spasial pada salah satu nomor (msisdn).

Tabel 4.29 Lokasi Home msisdn00xxx Untuk Setiap Bulan

event_month	msisdn	provider	home_prov	home_kab
12	msisdn00xxx	Telkomsel	JAWA TIMUR	LUMAJANG
1	msisdn00xxx	Telkomsel	JAWA TIMUR	LUMAJANG
2	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
3	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
4	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
5	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
6	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
7	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
8	msisdn00xxx	Telkomsel	JAWA TIMUR	JEMBER
9	msisdn00xxx	Telkomsel	DKI JAKARTA	JAKARTA TIMUR
10	msisdn00xxx	Telkomsel	DKI JAKARTA	JAKARTA TIMUR
11	msisdn00xxx	Telkomsel	DKI JAKARTA	JAKARTA TIMUR

Berdasarkan Tabel 4.29, sesuai dengan penentuan lokasi *home yearly* seperti yang dijelaskan sebelumnya, maka msisdn00xxx memiliki lokasi *home yearly* di Provinsi Jawa Timur, Kabupaten Jember. Oleh karena itu ekstraksi lintasan dimulai dari *record* pertama yang menunjukkan *event* dengan lokasi *home* berada di Kabupaten Jember (Februari 2022) sampai dengan *record* terakhir yang menunjukkan event dengan lokasi *home* yang sama, yakni Kabupaten Jember (Agustus 2022). Penentuan lokasi *start* dan *finish* dalam ekstraksi lintasan bertujuan agar nomor-nomor yang berada di kelompok yang sama (memiliki lokasi *home yearly* yang sama) memiliki titik *start* dan *finish* yang sama. Hal ini sesuai dengan pendekatan dalam GPS (*MPD active*) *trajectories clustering*.

dimana lokasi *start* dan *finish* untuk setiap lintasan yang akan dianalisis harus sama.

Setelah dilakukan ekstraksi lintasan yang terdiri dari *staypoint* terurut berdasarkan datetime, langkah selanjutnya adalah menghitung jarak antar lintasan dalam kelompok yang sama menggunakan DTW dan *Frechet Distance*. Hasil penghitungan tersebut disimpan dalam sebuah dataset untuk nantinya akan dibangun threshold menggunakan *gini impurity*.

Tabel 4.30 Contoh Jarak DTW antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Double	Jarak_DTW
msisdn00xxx06	msisdn00xxx05	TRUE	56073.94
msisdn00xxx07	msisdn00xxx05	FALSE	363313.9
msisdn00xxx07	msisdn00xxx06	FALSE	327057.7
msisdn00xxx08	msisdn00xxx05	FALSE	351945.7
msisdn00xxx08	msisdn00xxx06	FALSE	242016.3
msisdn00xxx08	msisdn00xxx07	TRUE	148682

Tabel 4.31 Contoh Jarak Frechet antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Double	Jarak_Frechet
msisdn00xxx06	msisdn00xxx05	TRUE	245.37
msisdn00xxx07	msisdn00xxx05	FALSE	1050.63
msisdn00xxx07	msisdn00xxx06	FALSE	1051.09
msisdn00xxx08	msisdn00xxx05	FALSE	1050.63
msisdn00xxx08	msisdn00xxx06	FALSE	1051.09



msisdn00xxx08	msisdn00xxx07	TRUE	643.49
---------------	---------------	------	--------

4.4.10 Pembangunan *Threshold*

Pembangunan *threshold* dilakukan dengan memanfaatkan *gini impurity* yang merupakan salah satu metode untuk menentukan *split node* terbaik di dalam *decision tree*. Pembangunan *threshold* dilakukan pada semua metode baik untuk algoritma *family grouping* maupun *double counting*. Variabel independen yang digunakan dalam penghitungan *threshold* ini adalah rata-rata jarak kestabilan antar nomor pada metode kombinasi jarak terdekat antar *staypoint*, jumlah irisan heksagon pada metode H3 tanpa *staypoint*, rata-rata jumlah irisan heksagon pada metode H3 dengan *staypoint*, jarak DTW dan jarak Frechet antar lintasan pada metode DTW dan *Frechet Distance*. Sedangkan kelas target atau variabel dependen yang digunakan adalah kelas *actual family grouping* dan *double counting*. Berikut akan dicontohkan penentuan *threshold* untuk entitas *family grouping* pada metode kombinasi jarak terdekat antar *staypoint*.

Tabel 4.32 Contoh Dataset Rata-rata Kestabilan Jarak Terdekat antar Nomor (MSISDN)

MSISDN 1	MSISDN 2	Is_Family	Average_D (km)
msisdn00xxx92	msisdn00xxx109	TRUE	0.015145
msisdn00xxx42	msisdn00xxx79	FALSE	9.926563
msisdn00xxx79	msisdn00xxx101	FALSE	8.224279
mslsdn00xxx42	mslsdn00xxx101	TRUE	0.000000
msisdn00xxx51	msisdn00xxx100	TRUE	0.140389
msisdn00xxx35	msisdn00xxx98	TRUE	0.140924
msisdn00xxx18	msisdn00xxx21	TRUE	0.395004



msisdn00xxx49	msisdn00xxx111	TRUE	2.445019
msisdn00xxx110	msisdn00xxx111	TRUE	2.113352
mslsdn00xxx91	mslsdn00xxx111	TRUE	0.744306

Langkah-langkah penentuan *threshold* (ambang batas) menggunakan *gini impurity* secara singkat adalah sebagai berikut:

1. Urutkan nilai variabel Average_D dari yang terkecil hingga yang terbesar.
2. Identifikasi titik-titik pemisahan potensial di antara nilai-nilai variabel Average_D yang berurutan menggunakan rata-rata sebagai kandidat pemisahan.
3. Untuk setiap titik pemisahan potensial, pisahkan nilai variabel Average_D menjadi dua kelompok berdasarkan apakah nilai variabel Average_D lebih kecil atau lebih besar dari titik pemisahan tersebut.
4. Hitung gini impurity untuk setiap pemisahan potensial.
5. Pilih pemisahan yang menghasilkan penurunan gini impurity yang paling signifikan sebagai pemisahan optimal untuk variabel Average_D tersebut.

Tabel 4.33 merupakan tabel hasil pengurutan baris berdasarkan nilai variabel Average_D dari yang terkecil hingga yang terbesar dan penghitungan rata-rata antara variabel Average_D yang berurutan.



Tabel 4.33 Pengurutan Berdasarkan Nilai Average_D dan Hasil Penghitungan rata-rata Average_D

MSISDN 1	MSISDN 2	Is_Family	Average_D (km)	Rata-rata Average_D Berurutan (km)
mslsdn00xxx42	mslsdn00xxx101	TRUE	0.000000	
msisdn00xxx92	msisdn00xxx109	TRUE	0.015145	0.0075725
,msisdn00xxx51	msisdn00xxx100	TRUE	0.140389	0.0777670
msisdn00xxx35	msisdn00xxx98	TRUE	0.140924	0.1406565
msisdn00xxx18	msisdn00xxx21	TRUE	0.395004	0.2679640
mslsdn00xxx91	mslsdn00xxx111	TRUE	0.744306	0.5696550
msisdn00xxx110	msisdn00xxx111	TRUE	2.113352	1.4288290
msisdn00xxx49	msisdn00xxx111	TRUE	2.445019	2.2791855
msisdn00xxx79	msisdn00xxx101	FALSE	8.224279	5.3346490
msisdn00xxx42	msisdn00xxx79	FALSE	9.926563	9.0754210

Setelah diurutkan dan dihitung rata-rata untuk setiap variabel Average_D yang berurutan, langkah selanjutnya adalah menghitung *gini impurity* untuk setiap rata-rata Average_D (*/threshold*). Misal, *threshold* yang digunakan untuk melakukan *split* menjadi dua kelompok adalah rata-rata Average_D = 0.0075725, maka *gini impurity* dapat dihitung dengan cara berikut:

Kelas ≥ 0.0075725

1. Untuk kelas ≥ 0.0075725 dan TRUE, rasionya adalah $7/9 = 0.778$
2. Untuk kelas ≥ 0.0075725 dan FALSE, rasionya adalah $2/9 = 0.222$

Maka untuk kelas ≥ 0.0075725 , *impurity* dapat dihitung dengan formula berikut:

$$\text{Impurity} = 1 - (\text{probability of "TRUE"})^2 - (\text{probability of "FALSE"})^2$$

$$\text{Impurity} = 1 - (0.778)^2 - (0.222)^2 = 0.345679$$

Kelas < 0.0075725

1. Untuk kelas < 0.0075725 dan TRUE, rasionya adalah $1/1 = 1$
2. Untuk kelas < 0.0075725 dan FALSE, rasionya adalah $0/1 = 0$

Maka untuk kelas < 0.0075725 , *impurity* dapat dihitung sebagai berikut.

$$\text{Impurity} = 1 - (\text{probability of "TRUE"})^2 - (\text{probability of "FALSE"})^2$$

$$\text{Impurity} = 1 - (1)^2 - (0)^2 = 0$$

Gini impurity untuk *threshold* rata-rata *Average_D* = 0.0075725 adalah:

$$\text{Gini Impurity} = (0.345679 \times (9/10)) + (0 \times (1/10)) = 0.3111$$

Hasil penghitungan *gini impurity* untuk setiap *threshold* rata-rata *Average_D* yang berurutan dapat dilihat pada tabel 4.34.

Tabel 4.34 Gini Impurity untuk Setiap Threshold Rata-rata Average_D

MSISDN 1	MSISDN 2	Is_Family	Average_D (km)	Rata-rata Average_D	Gini Impurity Berurutan (km)
mslsdn00xxx42	mslsdn00xxx101	TRUE	0.000000		
msisdn00xxx92	msisdn00xxx109	TRUE	0.015145	0.0075725	0.311111
,msisdn00xxx51	msisdn00xxx100	TRUE	0.140389	0.0777670	0.3
msisdn00xxx35	msisdn00xxx98	TRUE	0.140924	0.1406565	0.285714
msisdn00xxx18	msisdn00xxx21	TRUE	0.395004	0.2679640	0.266667
mslsdn00xxx91	mslsdn00xxx111	TRUE	0.744306	0.5696550	0.24



msisdn00xxx110	msisdn00xxx111	TRUE	2.113352	1.4288290	0.2
msisdn00xxx49	msisdn00xxx111	TRUE	2.445019	2.2791855	0.133333
msisdn00xxx79	msisdn00xxx101	FALSE	8.224279	5.3346490	0
msisdn00xxx42	msisdn00xxx79	FALSE	9.926563	9.0754210	0.1778

Threshold terbaik adalah *threshold* yang memiliki nilai *gini impurity* paling kecil. Pada Tabel 4.34, terlihat bahwa *threshold* 5.3346490 memiliki nilai *gini impurity* yang paling kecil. Berdasarkan hal tersebut, ditentukan *threshold* rata-rata kestabilan jarak terdekat antar nomor adalah < 5.3346490 untuk kelas True FG (entitas *family grouping*) dan rata-rata kestabilan jarak terdekat antar nomor ≥ 5.3346490 untuk kelas yang diprediksi False FG (bukan entitas *family grouping*).

Langkah penentuan *threshold* tersebut juga diterapkan pada metode-metode lainnya dengan menggunakan variabel independennya masing-masing dengan sedikit perbedaan pada tahap penentuan akhir dalam memprediksi kelas True (entitas *family grouping* atau *double counting*) dan kelas False (bukan entitas *family grouping* atau *double counting*). Pada DTW dan *frechet distance*, kelas True terdiri dari pasangan nomor yang memiliki jarak kurang dari *threshold* yang sudah ditentukan dan kelas False terdiri dari pasangan nomor yang memiliki jarak lebih dari *threshold* yang sudah ditentukan. Sementara itu, pada metode H3, kelas True terdiri dari pasangan nomor yang memiliki jumlah irisan atau rata-rata jumlah irisan lebih dari *threshold* yang sudah ditentukan dan berlaku hal sebaliknya untuk kelas False.

Setelah melakukan prediksi entitas mana yang termasuk dan yang tidak termasuk ke dalam *family grouping* atau *double counting*, langkah selanjutnya adalah melakukan evaluasi dengan membuat *confusion matrix* dan menghitung ukuran evaluasi seperti *precision*, *recall*, akurasi, dan *f1-score*. Untuk mengevaluasi kinerja algoritma *family grouping* dan *double counting*, pengukuran lebih diprioritaskan terhadap nilai *f1-score*.



dari kelas True, yaitu kelas yang berisikan *subscribers* yang diprediksi sebagai entitas *family grouping* atau *double counting*.

4.5 Hasil Implementasi Algoritma

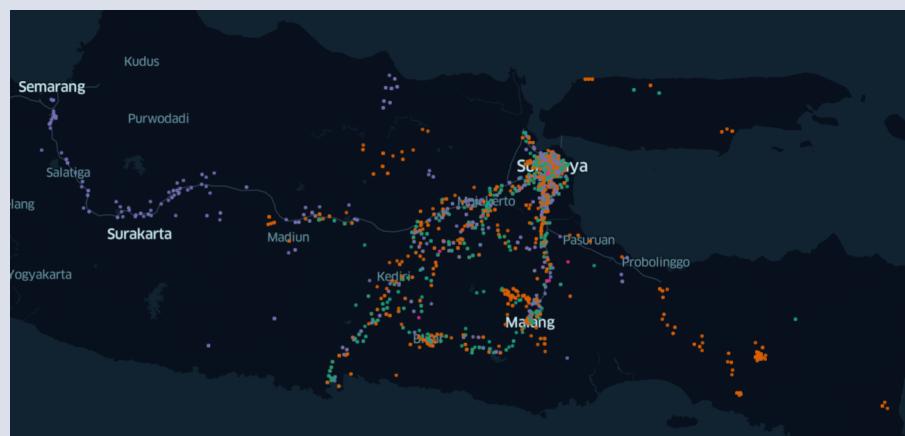
Algoritma *family grouping* dan *double counting* yang telah dijelaskan pada bagian sebelumnya diterapkan kepada *data training* untuk membangun *threshold* di masing-masing metode. *Threshold* yang telah dibangun selanjutnya digunakan untuk mengklasifikasikan *subscribers* pada *data testing* yang tergolong sebagai entitas *family grouping* atau *double counting* dan *subscribers* yang tidak tergolong sebagai entitas *family grouping* atau *double counting*. Kemudian, dilakukan evaluasi *performance* hasil klasifikasi yang telah dilakukan seperti yang telah dijelaskan di atas. Hasil evaluasi tiap metode akan dibandingkan guna mencari metode terbaik, yakni metode dengan ukuran evaluasi *performance* yang paling baik.

4.5.1 Family Grouping

4.5.1.1 Sebaran Raw Data

Untuk mengetahui bagaimana entitas *family grouping* dapat diidentifikasi secara umum, peneliti membuat visualisasi sebaran *raw data* dari kombinasi kartu sim Telkomsel dan Indosat yang digunakan oleh *subscribers family grouping*. Visualisasi ini bertujuan untuk melihat seberapa mirip sebaran *raw data* antar kombinasi kartu sim yang merepresentasikan pola perjalanan dari *subscribers family grouping* tersebut. Visualisasi sebaran *raw data* ini mengambil tiga sampel *family* secara acak, sampel pertama terdiri dari *family* yang menggunakan kartu sim Telkomsel seluruhnya, sampel kedua terdiri dari *family* yang menggunakan kartu sim Indosat seluruhnya, dan sampel ketiga terdiri dari *family* yang menggunakan kartu sim Telkomsel dan Indosat.

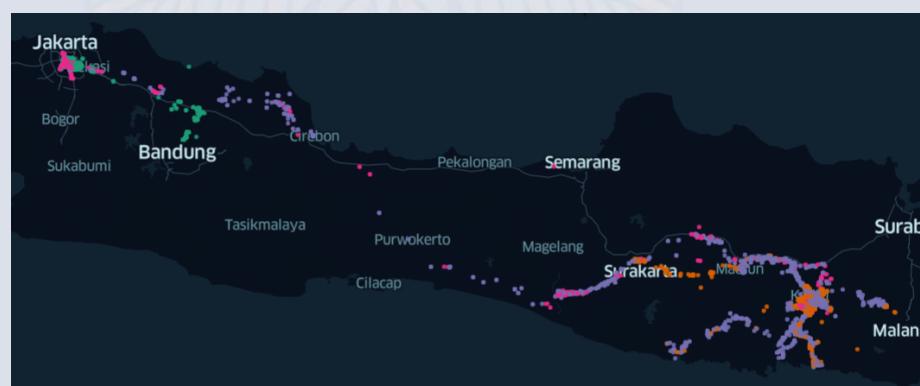
Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu SIM Telkomsel Seluruhnya



Gambar 4.55 Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu SIM Telkomsel Seluruhnya

Secara keseluruhan, sebaran *raw data* dari *family* yang menggunakan kartu sim Telkomsel seluruhnya tidak terlalu mirip. Berdasarkan Gambar 4.55, terlihat bahwa terdapat sebaran *raw data* dari dua *subscribers* yang memiliki pola perjalanan berbeda dengan *subscribers* lainnya. *Subscribers* tersebut adalah *subscribers* dengan sebaran *raw data* berwarna ungu dan oranye.

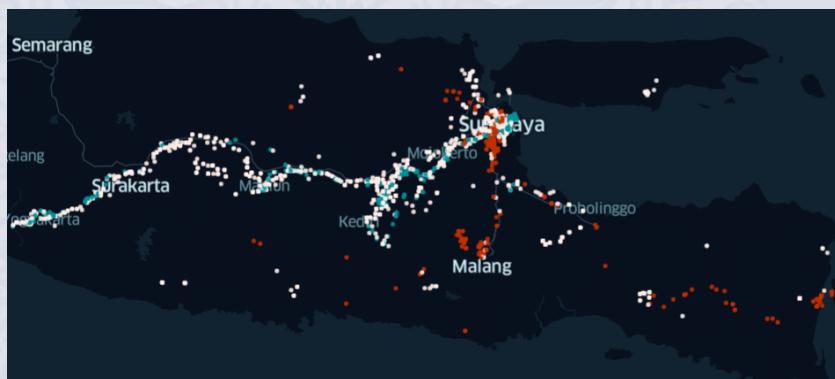
Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu sim Indosat Seluruhnya



Gambar 4.56 Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu SIM Indosat Seluruhnya

Tidak jauh berbeda dengan sebaran *raw data* dari *family* yang menggunakan kartu sim Telkomsel seluruhnya, *family* yang seluruh anggotanya menggunakan kartu sim Indosat juga memiliki sebaran *raw data* yang tidak terlalu mirip. Berdasarkan Gambar 4.56, dapat dilihat bahwa terdapat anggota keluarga yang seolah-olah tidak tinggal bersama di suatu tempat. Hal tersebut disebabkan oleh sebaran *raw data* berwarna merah muda dan hijau muda cenderung berada di daerah antara Jakarta sampai Bandung saja. Sedangkan sebaran *raw data* berwarna ungu dan oranye cenderung berada di daerah Surakarta dan beberapa daerah di Jawa Timur.

Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu sim Telkomsel dan Indosat



Gambar 4.57 Sebaran *Raw Data* dari *Family* yang Menggunakan Kartu SIM Telkomsel dan Indosat

Sama seperti sebaran *raw data* dari dua sampel *family* sebelumnya, sebaran *raw data* dari *family* yang menggunakan kartu sim Telkomsel dan Indosat juga tidak terlalu mirip. Berdasarkan Gambar 4.57, terlihat bahwa terdapat sebaran *raw data* dari salah satu *subscriber* yang memiliki pola perjalanan berbeda dengan *subscribers* lainnya. *Subscriber* tersebut adalah *subscriber* dengan sebaran *raw data* berwarna merah.



Dari hasil visualisasi ketiga sampel *family* yang menggunakan kombinasi kartu sim berbeda, dapat disimpulkan bahwa baik *family* yang menggunakan kartu sim Telkomsel seluruhnya, Indosat seluruhnya, maupun menggunakan Telkomsel dan Indosat, semuanya memiliki pola perjalanan yang cukup berbeda antar anggota keluarga, sehingga menyulitkan proses pengidentifikasi *subscribers* yang termasuk ke dalam anggota keluarga tertentu karena adanya ketidakmiripan pola perjalanan.

4.5.1.2 Evaluasi Performa Algoritma

Dari seluruh metode yang digunakan dalam pembangunan algoritma *family grouping*, pada subbab ini akan dilakukan perbandingan performa dari masing-masing metode dalam memprediksi entitas *family grouping* pada *data testing* sesuai dengan kelas aktualnya. Sebelum melakukan perbandingan performa antar metode, perlu diketahui terlebih dahulu bagaimana masing-masing metode dapat diimplementasikan ke dalam *data training* dengan baik dan juga bagaimana *threshold* yang telah dibangun dapat memisahkan kelas True FG dan kelas False FG dengan jelas.

A. Metode Kombinasi Jarak Terdekat antar *Staypoint*

Berikut ini merupakan perbandingan rata-rata jarak kestabilan antar kelas True DC dan kelas False DC dengan menggunakan metode kombinasi jarak terdekat antar *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jarak Kestabilan Antar Kelas pada Data Training Algoritma Family Grouping

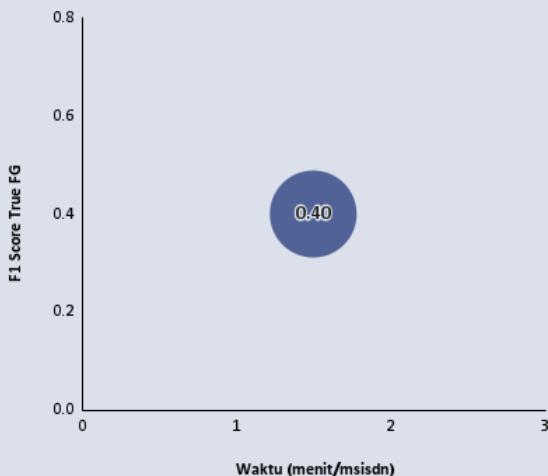


Gambar 4.58 Threshold dan Perbandingan Rata-Rata Jarak Kestabilan Antar Kelas pada Data Training Algoritma Family Grouping

Berdasarkan Gambar 4.58, terlihat bahwa algoritma *family grouping* dengan metode kombinasi jarak terdekat antar staypoint dapat menghasilkan rata-rata jarak kestabilan yang cukup berbeda antara kelas True FG dan False FG pada *data training*. Rata-rata jarak kestabilan untuk kelas True FG lebih kecil dibandingkan dengan kelas False FG. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *family grouping* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan jarak antar *staypoint* yang lebih kecil dibandingkan dengan bukan entitas *family grouping*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode kombinasi jarak terdekat antar *staypoint* dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family grouping*.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* kurang dapat memisahkan entitas *family grouping* dan bukan entitas *family grouping*. Hal tersebut ditandai dengan *threshold* yang memiliki nilai kurang dari rata-rata jarak kestabilan kelas True FG dan kelas False FG (bukan berada diantaranya), sehingga kurang baik dalam memisahkan entitas *family grouping* dan bukan entitas *family grouping* secara jelas.

**Hasil Implementasi Metode Kombinasi Jarak Terdekat Antar *Staypoint*
pada *Data Testing* Algoritma *Family Grouping***



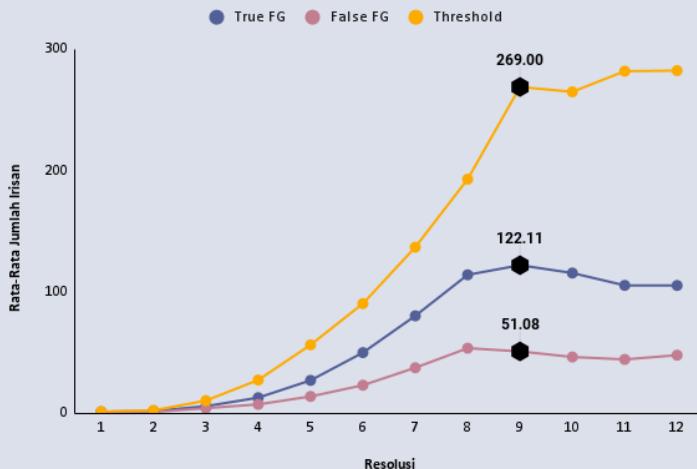
Gambar 4.59 Hasil Implementasi Metode Kombinasi Jarak Terdekat antar *Staypoint* pada *Data Testing* Algoritma *Family Grouping*

Hasil implementasi *threshold* yang telah dibangun terhadap *data testing* dapat dilihat pada Gambar 4.59. Berdasarkan gambar 4.59, dapat diketahui bahwa metode kombinasi jarak terdekat antar *staypoint* memiliki F1-Score kelas True DC sebesar 0,40 dengan rata-rata waktu *running* 1,5 menit untuk 1 msisdn/nomor. Dengan begitu, dapat disimpulkan bahwa metode kombinasi jarak terdekat antar *staypoint* ini kurang disarankan untuk digunakan dalam algoritma *family grouping* karena memiliki nilai F1-Score kelas True DC yang tergolong rendah (kurang akurat) dan waktu *running* yang cukup lama.

**B. Metode *Hexagonal Hierarchical Geospatial Indexing System* (H3)
Tanpa *Staypoint***

Berikut ini merupakan perbandingan jumlah irisan heksagon antar kelas True DC dan kelas False DC dengan menggunakan metode *hexagonal hierarchical geospatial indexing system* (H3) tanpa *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping



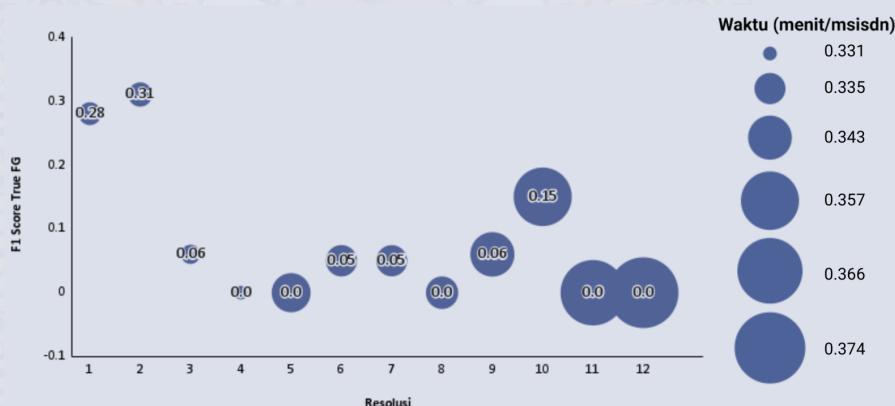
Gambar 4.60 Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping

Berdasarkan Gambar 4.60, terlihat bahwa algoritma *family grouping* dengan metode H3 tanpa *staypoint* dapat menghasilkan jumlah irisan heksagon yang cukup berbeda antara kelas True FG dan False FG pada *data training*, terutama dari resolusi 5 hingga resolusi 12, dengan jumlah irisan heksagon terbesar pada resolusi 9. Hal ini menunjukkan bahwa metode H3 tanpa *staypoint* dengan resolusi 9 dapat dengan jelas memisahkan antara kelas True FG dan False FG daripada resolusi lainnya. Selain itu, secara keseluruhan, jumlah irisan heksagon untuk kelas True FG lebih besar dibandingkan dengan kelas False FG. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *family grouping* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan jumlah heksagon yang beririsan lebih banyak dibandingkan dengan bukan entitas *family grouping*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode H3 tanpa *staypoint* dapat secara jelas memisahkan antara entitas *family grouping* dan bukan entitas *family grouping*.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* kurang dapat memisahkan entitas *family grouping* dan bukan entitas *family grouping*. Hal tersebut ditandai dengan *threshold* yang memiliki nilai lebih besar dari jumlah irisan heksagon kelas True FG dan kelas False FG (bukan berada di antara kedua kelas) di seluruh resolusi sehingga *threshold* kurang baik dalam memisahkan entitas *family grouping* dan bukan entitas *family grouping* secara jelas.

Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing

Algoritma Family Grouping



Gambar 4.61 Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing Algoritma Family Grouping

Hasil implementasi *threshold* yang telah dibangun terhadap *data testing* dapat dilihat pada Gambar 4.61. Berdasarkan Gambar 4.61, terlihat bahwa metode H3 tanpa *staypoint* memiliki F1-Score kelas True FG tertinggi pada resolusi 2 yaitu sebesar 0,31. Selain itu, resolusi 2 memiliki waktu *running* yang cukup singkat dibandingkan dengan resolusi lainnya (resolusi 5-12) yaitu hanya selama 0,33 menit atau sekitar 20 detik untuk 1 msisdn/nomor. Terlihat juga bahwa semakin tinggi resolusi H3, waktu *running* juga semakin lama. Hal tersebut disebabkan oleh semakin banyaknya *cell heksagon* hasil konversi dari *staypoint* seiring dengan semakin tingginya resolusi H3 sehingga proses perbandingan untuk

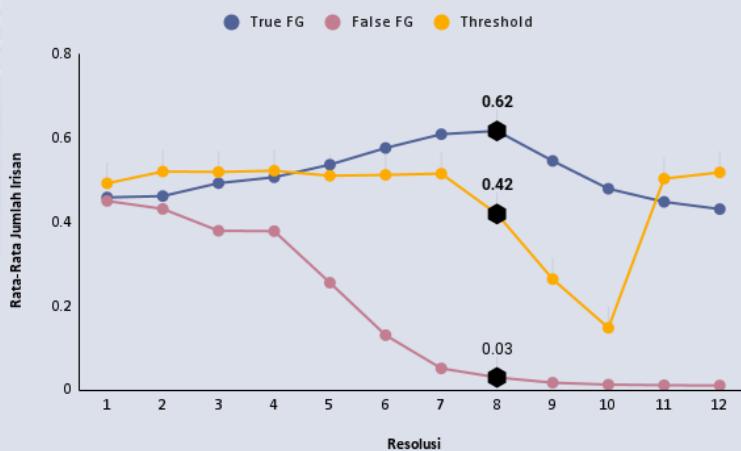
mencari heksagon yang beririsan antar dua nomor juga semakin kompleks dan membutuhkan waktu yang tidak sebentar.

Berdasarkan hasil dan penjelasan di atas, dapat disimpulkan bahwa resolusi 2 adalah resolusi terbaik pada metode H3 tanpa *staypoint*. Namun metode ini masih memiliki nilai F1-Score kelas True FG yang tergolong rendah (kurang akurat) sehingga kurang disarankan untuk digunakan dalam algoritma *family grouping*.

C. Metode *Hexagonal Hierarchical Geospatial Indexing System* (H3) dengan *Staypoint*

Berikut ini merupakan perbandingan rata-rata jumlah irisan heksagon antar kelas True DC dan kelas False DC dengan menggunakan metode *hexagonal hierarchical geospatial indexing system* (H3) dengan *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping



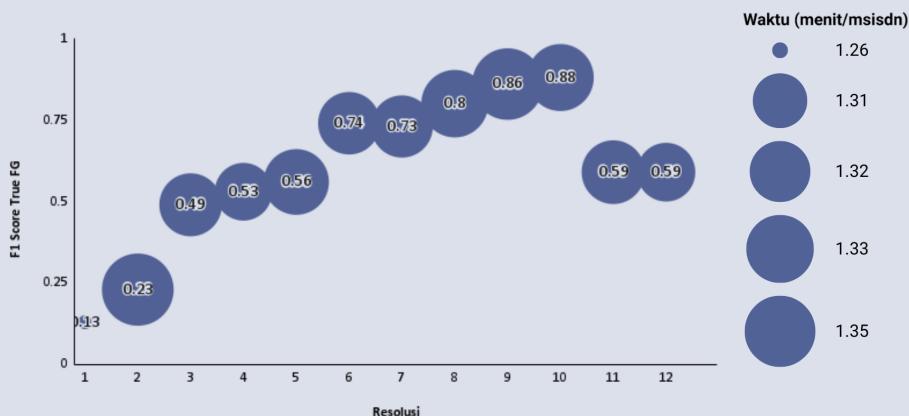
Gambar 4.62 Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Family Grouping



Berdasarkan Gambar 4.62, terlihat bahwa algoritma *family grouping* dengan metode H3 dengan *staypoint* dapat menghasilkan rata-rata jumlah irisan heksagon yang cukup berbeda antara kelas True FG dan False FG pada *data training*, terutama dari resolusi 3 hingga resolusi 12, dengan selisih rata-rata jumlah irisan heksagon terbesar pada resolusi 8. Hal ini menunjukkan bahwa metode H3 dengan *staypoint* dengan resolusi 8 dapat dengan jelas memisahkan kelas True FG dan False FG dibandingkan dengan resolusi lainnya. Secara keseluruhan, rata-rata jumlah irisan heksagon untuk kelas True FG lebih besar dibandingkan dengan kelas False FG. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *family grouping* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jumlah heksagon yang beririsan per hari lebih banyak dibandingkan dengan bukan entitas *family grouping*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode H3 dengan *staypoint* dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family grouping* terutama pada resolusi 3 hingga resolusi 12.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family grouping* pada resolusi 5 hingga resolusi 10. Hal tersebut ditandai dengan nilai *threshold* yang berada diantara rata-rata jumlah irisan heksagon kelas True FG dan kelas False FG. Sementara itu, *threshold* pada resolusi 1, 2, 3, 4, 11 dan 12 kurang dapat memisahkan antara entitas kedua kelas. Hal tersebut dapat terlihat dari *threshold* yang bernilai lebih dari rata-rata jumlah irisan heksagon kelas True FG dan kelas False FG.

Hasil Implementasi Metode H3 dengan *Staypoint* pada *Data Testing* Algoritma *Family Grouping*



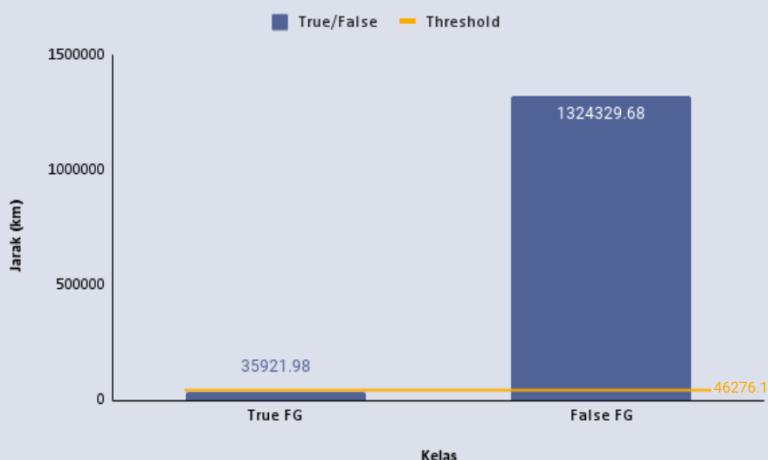
Gambar 4.63 Hasil Implementasi Metode H3 dengan *Staypoint* pada *Data Testing* Algoritma *Family Grouping*

Hasil implementasi *threshold* yang telah dibangun terhadap *data testing* dapat dilihat pada Gambar 4.63. Berdasarkan gambar 4.63, dapat diketahui bahwa metode H3 dengan staypoint memiliki F1-Score kelas True FG tertinggi pada resolusi 10 yaitu sebesar 0,88. Selain itu resolusi 10 juga memiliki waktu *running* yang sedikit lebih singkat dibandingkan dengan resolusi lainnya yaitu selama 1,32 menit untuk 1 msisdn/nomor. Dari grafik juga dapat dilihat bahwa tidak terdapat perbedaan waktu *running* yang cukup signifikan antar resolusi. Berdasarkan hasil tersebut, dapat disimpulkan bahwa resolusi 10 adalah resolusi terbaik pada metode H3 dengan *staypoint* karena tergolong sangat akurat dalam mengidentifikasi entitas *family grouping* dan memiliki waktu *running* yang sedikit lebih singkat dibandingkan dengan resolusi lainnya.

D. Metode *Dynamic Time Warping* (DTW)

Berikut ini merupakan perbandingan rata-rata jarak DTW antar kelas True DC dan kelas False DC beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jarak DTW Antar Kelas pada Data Training Algoritma Family Grouping



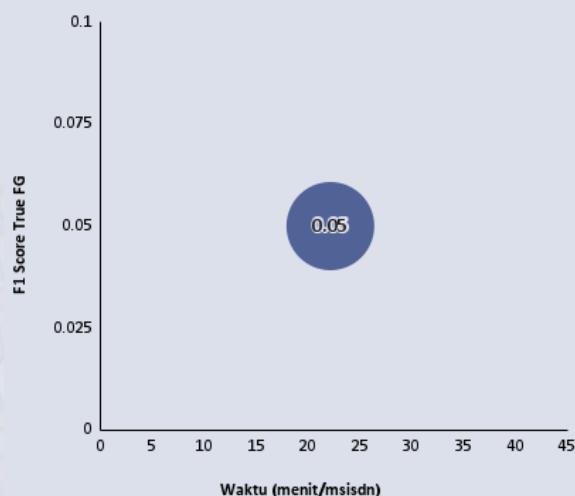
Gambar 4.64 Threshold dan Perbandingan Rata-rata Jarak DTW Antar Kelas pada Data Training Algoritma Family Grouping

Berdasarkan Gambar 4.64, dapat dilihat bahwa algoritma *family grouping* dengan metode DTW dapat menghasilkan rata-rata jarak DTW yang cukup berbeda antara kelas True FG dan False FG pada *data training*. Rata-rata jarak DTW untuk kelas True FG lebih kecil dibandingkan dengan kelas False FG. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *family grouping* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jarak DTW yang lebih kecil dibandingkan dengan bukan entitas *family grouping*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode DTW dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family grouping*.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family*

grouping. Hal tersebut ditandai dengan *threshold* yang memiliki nilai berada diantara rata-rata jarak DTW kelas True FG dan kelas False FG.

Hasil Implementasi Metode DTW pada *Data Testing Algoritma Family Grouping*



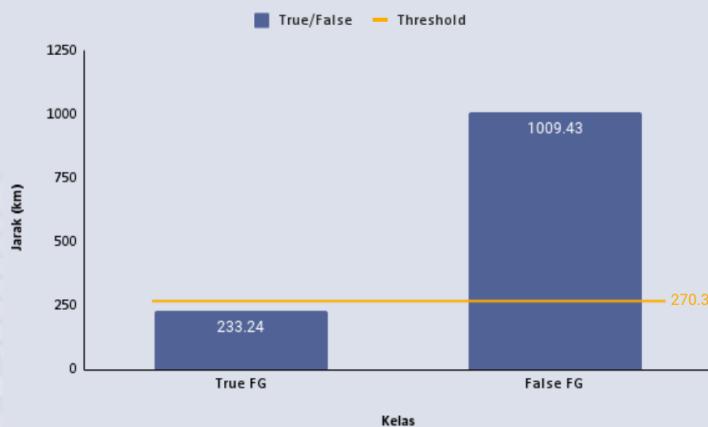
Gambar 4.65 Hasil Implementasi Metode DTW pada *Data Testing Algoritma Family Grouping*

Hasil implementasi *threshold* yang telah dibangun terhadap *data testing* dapat dilihat pada Gambar 4.65. Berdasarkan gambar 4.65, dapat diketahui bahwa metode DTW memiliki F1-Score kelas True FG hanya sebesar 0,05 dengan rata-rata waktu *running* sekitar 22 menit untuk 1 msisdn/nomor. Dengan begitu, dapat disimpulkan bahwa metode DTW ini tidak disarankan untuk digunakan dalam algoritma *family grouping* karena memiliki nilai F1-Score kelas True FG yang tergolong sangat rendah (tidak akurat) dan waktu running yang sangat lama.

E. Metode *Frechet Distance*

Berikut ini merupakan perbandingan rata-rata jarak Frechet antar kelas True DC dan kelas False DC beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Family Grouping



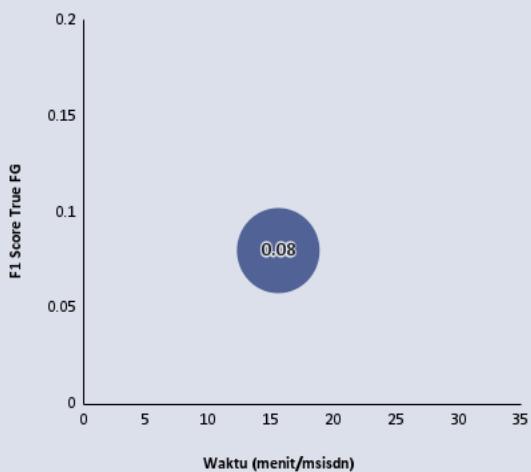
Gambar 4.66 Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Family Grouping

Berdasarkan Gambar 4.66, dapat dilihat bahwa algoritma *family grouping* dengan metode Frechet Distance dapat menghasilkan rata-rata jarak Frechet yang cukup berbeda antara kelas True FG dan False FG pada *data training*. Rata-rata jarak Frechet untuk kelas True FG lebih kecil dibandingkan dengan kelas False FG. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *family grouping* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jarak Frechet yang lebih kecil dibandingkan dengan bukan entitas *family grouping*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode Frechet Distance dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family grouping*.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* dapat secara jelas memisahkan entitas *family grouping* dan bukan entitas *family*

grouping. Hal tersebut ditandai dengan *threshold* yang memiliki nilai berada diantara rata-rata jarak Frechet kelas True FG dan kelas False FG.

Hasil Implementasi Metode *Frechet Distance* pada *Data Testing* Algoritma *Family Grouping*



Gambar 4.67 Hasil Implementasi Metode *Frechet Distance* pada *Data Testing* Algoritma *Family Grouping*

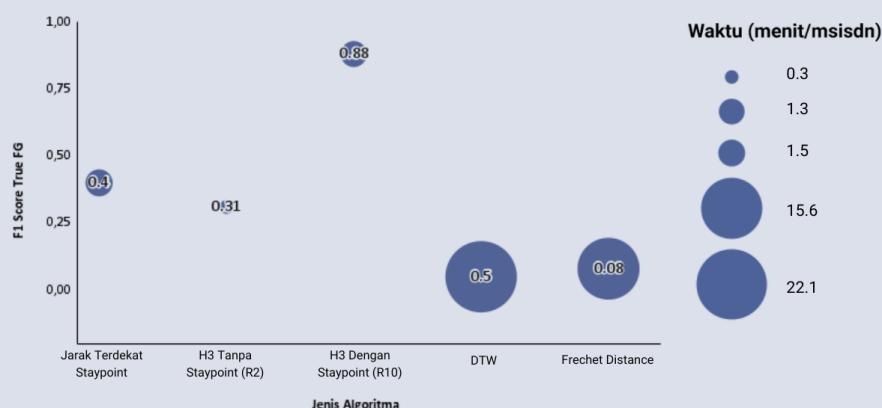
Hasil implementasi *threshold* yang telah dibangun terhadap *data testing* dapat dilihat pada Gambar 4.67. Berdasarkan gambar 4.67, dapat diketahui bahwa metode *Frechet Distance* memiliki F1-Score kelas True FG hanya sebesar 0,08 dengan rata-rata waktu *running* sekitar 16 menit untuk 1 msisdn/nomor. Dengan begitu, dapat disimpulkan bahwa metode Frechet Distance ini tidak disarankan untuk digunakan dalam algoritma *family grouping* karena memiliki nilai F1-Score kelas True FG yang tergolong sangat rendah (tidak akurat) dan waktu running yang sangat lama.

4.5.1.3 Perbandingan Performa Algoritma Terbaik

Algoritma terbaik dari masing-masing metode kemudian dibandingkan satu sama lain untuk menentukan algoritma yang paling baik dalam mengklasifikasikan entitas *family grouping* sesuai dengan kelas

aktualnya pada *data testing*. Hasil perbandingan dapat terlihat pada gambar di bawah ini.

Perbandingan Algoritma *Family Grouping* Terbaik dari Masing-masing Metode



Gambar 4.68 Perbandingan Algoritma *Family Grouping* Terbaik dari Masing-masing Metode

Berdasarkan Gambar 4.68, terlihat bahwa algoritma terbaik yang dapat digunakan untuk menangani fenomena *family grouping* adalah algoritma H3 dengan *staypoint* resolusi 10. Algoritma ini memiliki nilai F1-Score kelas True FG yang paling tinggi diantara algoritma lainnya meskipun waktu *running* dari algoritma ini tergolong cukup lama, yakni 1,32 menit untuk 1 msisdn/nomor. Namun, waktu *running* tersebut masih lebih singkat daripada waktu *running* algoritma lainnya. Terdapat satu algoritma lainnya yang memiliki waktu *running* lebih singkat, yaitu algoritma H3 tanpa *staypoint* resolusi 2 dengan waktu *running* hanya selama 0,33 menit atau sekitar 20 detik untuk 1 msisdn/nomor, tetapi nilai F1-Score kelas True FG dari algoritma ini lebih kecil daripada algoritma H3 dengan *staypoint* resolusi 10. Sementara itu, algoritma *family grouping* dengan pendekatan DTW dan *Frechet Distance* memiliki performa yang buruk dalam mengidentifikasi entitas *family grouping* karena memiliki nilai F1-Score kelas True FG yang sangat rendah dan membutuhkan waktu

running yang sangat lama. Berikut adalah *classification report* dari algoritma H3 dengan *staypoint* resolusi 10.

Tabel 4.35 Classification Report Algoritma H3 dengan Staypoint Resolusi 10

Kelas	Precision	Recall	F1-Score	Support
True DC	0,97	0,81	0,88	36
False DC	0,98	1,00	0,99	319
Macro Avg	0,97	0,90	0,93	355
Accuracy			0,98	

Berdasarkan Tabel 4.35, terlihat bahwa algoritma H3 dengan *staypoint* resolusi 10 memiliki nilai *macro average F1-Score* yang sangat tinggi yaitu sebesar 0,93. Secara keseluruhan, algoritma ini memiliki akurasi yang sangat tinggi juga yaitu sebesar 0,98 yang artinya algoritma ini dapat dengan tepat mengklasifikasikan 98% entitas *family grouping* dan bukan entitas *family grouping* sesuai dengan kelas aktualnya pada *data testing*.

4.5.2 Double Counting

4.5.2.1 Sebaran Raw Data

Untuk mengetahui bagaimana entitas *double counting* dapat diidentifikasi secara umum, peneliti membuat visualisasi sebaran *raw data* dari kombinasi kartu sim Telkomsel dan Indosat yang digunakan oleh *subscribers double counting*. Visualisasi ini bertujuan untuk melihat seberapa mirip sebaran *raw data* antar kombinasi kartu sim yang merepresentasikan pola perjalanan dari *subscriber double counting* tersebut. Visualisasi sebaran *raw data* ini mengambil tiga sampel secara acak, yakni sampel *subscriber double counting* dengan kombinasi kartu sim Telkomsel-Telkomsel, kombinasi kartu sim Indosat-Indosat, dan sampel kombinasi kartu sim Telkomsel-Indosat.

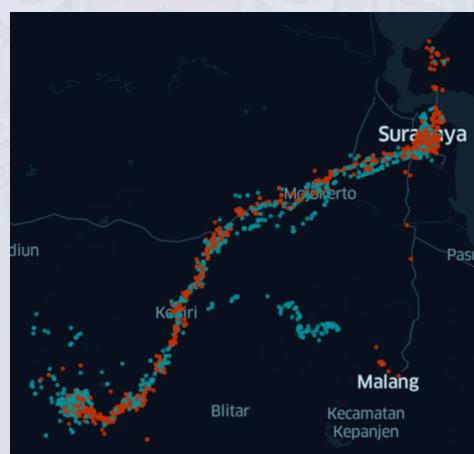
**Sebaran *Raw Data Subscriber Double Counting* dengan Kombinasi Kartu
SIM Telkomsel (Merah) - Telkomsel (Hijau)**



Gambar 4.69 Sebaran *Raw Data Subscriber Double Counting* dengan
Kombinasi Kartu SIM Telkomsel (Merah) - Telkomsel (Hijau)

Secara keseluruhan, sebaran *raw data* dari *subscriber double counting* yang menggunakan kombinasi kartu sim Telkomsel-Telkomsel relatif mirip. Hampir di setiap perjalanan *subscriber* tersebut, terdapat sebaran *raw data* dari kedua kartu sim.

**Sebaran *Raw Data Subscriber Double Counting* dengan Kombinasi Kartu
SIM Indosat (Merah) - Indosat (Hijau)**



Gambar 4.70 Sebaran *Raw Data Subscriber Double Counting* dengan
Kombinasi Kartu SIM Indosat (Merah) - Indosat (Hijau)

Tidak jauh berbeda dengan sebaran *raw data* kombinasi kartu sim Telkomsel-Telkomsel, *subscriber double counting* dengan kombinasi kartu sim Indosat-Indosat juga memiliki sebaran *raw data* yang relatif mirip. Berdasarkan Gambar 4.70, terlihat bahwa *subscriber double counting* tersebut, berdasarkan catatan lokasi pada data MPD, sering melakukan mobilitas dari Kabupaten Trenggalek menuju Kota Surabaya dan sebaliknya yang mana di setiap perjalanannya terdapat sebaran *raw data* dari kedua kartu sim.

Sebaran Raw Data Subscriber Double Counting dengan Kombinasi Kartu SIM Telkomsel (Merah) - Indosat (Hijau)



Gambar 4.71 Sebaran *Raw Data Subscriber Double Counting* dengan Kombinasi Kartu SIM Telkomsel (Merah) - Indosat (Hijau)

Berbeda dengan sebaran *raw data* kombinasi kartu sim Telkomsel-Telkomsel dan Indosat-Indosat, *subscriber double counting* dengan kombinasi kartu sim Telkomsel-Indosat memiliki sebaran *raw data* yang tidak terlalu mirip. Bahkan terdapat perjalanan *subscriber* yang hanya memiliki sebaran *raw data* dari kartu sim Telkomsel saja, tanpa ada sebaran *raw data* dari kartu sim Indosat. Padahal kedua kartu sim tersebut dipakai secara bersamaan dalam satu *handphone* utama.



Dari hasil visualisasi ketiga sampel *subscribers double counting*, dapat disimpulkan bahwa *subscribers* dengan kombinasi kartu sim Telkomsel-Telkomsel dan Indosat-Indosat memiliki pola perjalanan yang cukup mirip sehingga *subscribers* tersebut mudah untuk diidentifikasi sebagai entitas *double counting* berdasarkan kemiripan pola perjalanannya. Sementara itu, *subscriber double counting* dengan kombinasi kartu sim Telkomsel-Indosat memiliki sebaran *raw data* yang cukup berbeda dan hal ini membuat *subscriber* tersebut menjadi sulit untuk diidentifikasi sebagai entitas *double counting* karena adanya ketidakmiripan pola perjalanannya.

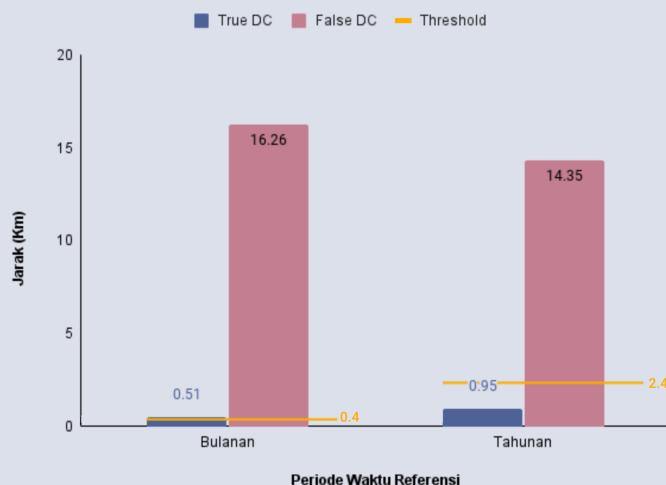
4.5.2.2 Evaluasi Performa Algoritma

Dari seluruh metode yang digunakan dalam pembangunan algoritma *double counting*, pada subbab ini, performa masing-masing metode dalam memprediksi entitas *double counting* pada *data testing* sesuai dengan kelas aktualnya akan dibandingkan. Sebelum membandingkan performa antar metode, perlu diketahui terlebih dahulu bagaimana masing-masing metode dapat diimplementasikan ke dalam *data training* dengan baik dan juga bagaimana *threshold* yang telah dibangun dapat memisahkan kelas True DC dan kelas False DC dengan jelas.

A. Metode Kombinasi Jarak Terdekat antar *Staypoint*

Berikut ini merupakan perbandingan rata-rata jarak kestabilan antar kelas True DC dan kelas False DC dengan menggunakan metode kombinasi jarak terdekat antar *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jarak Kestabilan antar Kelas pada Data Training Algoritma Double Counting



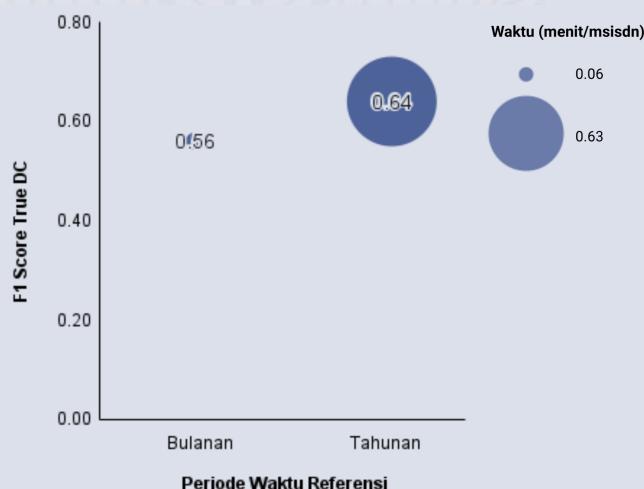
Gambar 4.72 Threshold dan Perbandingan Rata-rata Jarak Kestabilan antar Kelas pada Data Training Algoritma Double Counting

Berdasarkan Gambar 4.72, terlihat bahwa algoritma *double counting* dengan metode kombinasi jarak terdekat antar *staypoint* dapat menghasilkan rata-rata jarak kestabilan yang cukup berbeda antara kelas True DC dan False DC pada *data training*, baik pada periode waktu referensi bulanan maupun tahunan. Rata-rata jarak kestabilan untuk kelas True DC lebih kecil dibandingkan dengan kelas False DC. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *double counting* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan jarak antar *staypoint* yang lebih kecil dibandingkan dengan bukan entitas *double counting*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode kombinasi jarak terdekat antar *staypoint* dapat secara jelas memisahkan antara entitas *double counting* dan bukan entitas *double counting*.

Apabila dilihat berdasarkan *threshold* yang telah ditetapkan menggunakan *gini impurity*, dapat diketahui bahwa *threshold* dapat secara jelas memisahkan antara entitas *double counting* dan bukan entitas *double counting* pada periode waktu referensi tahunan. Hal tersebut ditandai dengan nilai *threshold* yang berada di antara rata-rata jarak kestabilan

kelas True DC dan kelas False DC. Berbeda dengan *threshold* pada periode waktu referensi bulanan yang memiliki nilai lebih kecil dari rata-rata jarak kestabilan kelas True DC (bukan berada diantaranya) sehingga *threshold* tersebut kurang baik dalam memisahkan antara kedua kelas entitas secara jelas. Dengan begitu, dapat disimpulkan bahwa *threshold* pada periode waktu referensi tahunan lebih baik dibandingkan dengan *threshold* pada periode waktu referensi bulanan.

**Hasil Implementasi Metode Kombinasi Jarak Terdekat antar *Staypoint* pada
Data Testing Algoritma Double Counting**



Gambar 4.73 Hasil Implementasi Metode Kombinasi Jarak Terdekat antar *Staypoint* pada *Data Testing* Algoritma Double Counting

Hasil implementasi *threshold* yang telah ditetapkan terhadap *data testing* terlihat pada Gambar 4.73. Berdasarkan gambar 4.73, dapat diketahui bahwa metode kombinasi jarak terdekat antar *staypoint* memiliki F1-Score kelas True DC pada periode waktu referensi tahunan lebih besar daripada periode waktu referensi bulanan, yakni 0,64 untuk tahunan dan 0,56 untuk bulanan. Namun, rata-rata waktu *running* dari periode waktu referensi tahunan juga jauh lebih besar daripada periode waktu referensi bulanan, yaitu sebesar 0,63 menit atau sekitar 38 detik untuk 1 msisdn/nomor dengan waktu referensi tahunan dan 0,06 menit atau sekitar 4 detik untuk 1 msisdn/nomor dengan waktu referensi bulanan. Rata-rata

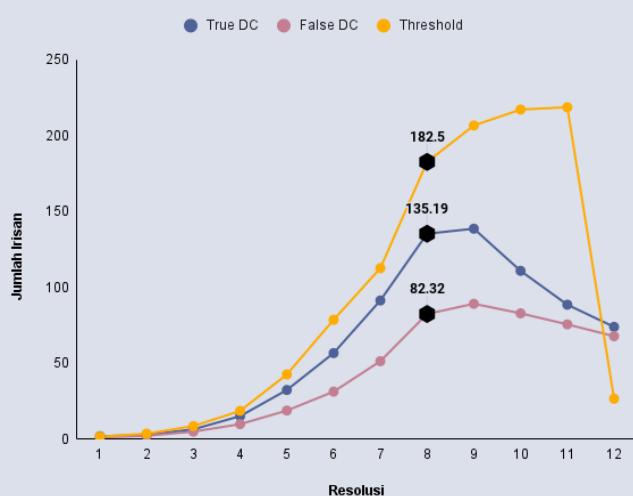
waktu *running* periode waktu referensi tahunan yang lebih lama disebabkan oleh rentang waktu yang dibandingkan dan dilakukan penghitungan antar nomor satu dengan nomor lainnya lebih lama dibandingkan dengan periode waktu referensi bulanan.

Dari penjelasan di atas, dapat disimpulkan bahwa jika peneliti ingin lebih memprioritaskan tingkat keakuratan dalam mengidentifikasi entitas *double counting*, metode kombinasi jarak terdekat antar *staypoint* dengan periode waktu referensi tahunan dapat digunakan. Namun, apabila peneliti ingin lebih memprioritaskan tingkat efisiensi waktu *running*, metode kombinasi jarak terdekat antar *staypoint* dengan periode waktu referensi bulanan dapat digunakan.

B. Metode *Hexagonal Hierarchical Geospatial Indexing System (H3)* *Tanpa Staypoint*

Berikut ini merupakan perbandingan jumlah irisan heksagon antar kelas True DC dan kelas False DC dengan metode H3 tanpa *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting



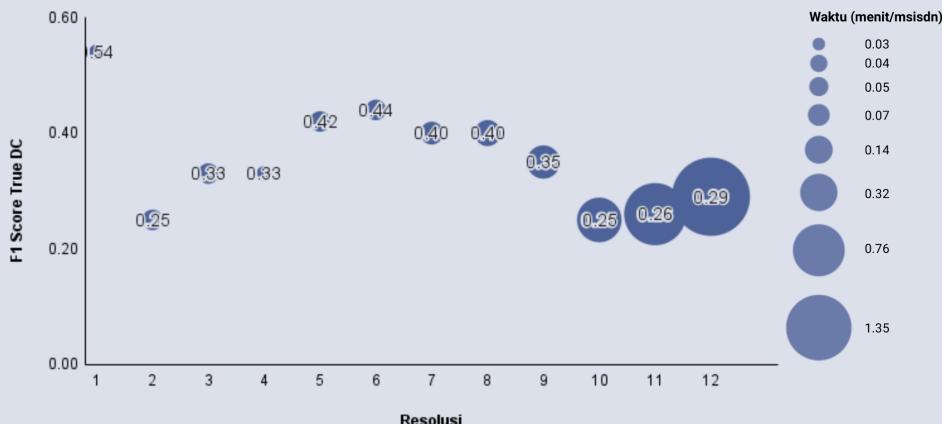
Gambar 4.74 Threshold dan Perbandingan Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting



Berdasarkan Gambar 4.74, terlihat bahwa algoritma *double counting* dengan metode H3 tanpa *staypoint* dapat menghasilkan jumlah irisan heksagon yang cukup berbeda antara kelas True DC dan False DC pada *data training*, terutama dari resolusi 5 hingga resolusi 11. Namun, diantara seluruh resolusi, resolusi 8 memiliki selisih jumlah irisan heksagon terbesar antar kelas. Hal ini menunjukkan bahwa metode H3 tanpa *staypoint* dengan resolusi 8 dapat dengan jelas memisahkan antara kelas True DC dan False DC dibandingkan dengan resolusi lainnya. Selain itu, secara keseluruhan, jumlah irisan heksagon untuk kelas True DC lebih besar dibandingkan dengan kelas False DC. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *double counting* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan jumlah heksagon yang beririsan lebih banyak dibandingkan dengan bukan entitas *double counting*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode H3 tanpa *staypoint* dapat secara jelas memisahkan antara entitas *double counting* dan bukan entitas *double counting*.

Apabila dilihat berdasarkan *threshold* yang telah ditetapkan menggunakan *gini impurity*, dapat diketahui bahwa *threshold* kurang dapat memisahkan entitas *double counting* dan bukan entitas *double counting*. Hal tersebut ditandai dengan *threshold* yang memiliki nilai lebih besar dari jumlah irisan heksagon kelas True DC dan kelas False DC (bukan berada di antara kedua kelas) di seluruh resolusi, kecuali resolusi 12, sehingga *threshold* kurang baik dalam memisahkan entitas *double counting* dan bukan entitas *double counting* secara jelas.

Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing Algoritma Double Counting



Gambar 4.75 Hasil Implementasi Metode H3 Tanpa Staypoint pada Data Testing
Algoritma Double Counting

Hasil implementasi *threshold* yang telah ditetapkan terhadap *data testing* dapat dilihat pada Gambar 4.75. Berdasarkan gambar 4.75, terlihat bahwa metode H3 tanpa *staypoint* memiliki F1-Score kelas True DC tertinggi pada resolusi 6 yaitu sebesar 0,44. Perlu diketahui bahwa terdapat pengecualian pada resolusi 1 karena, berdasarkan grafik, terdapat *trend* nilai F1-Score kelas True DC yang semakin kecil seiring dengan semakin kecilnya resolusi. oleh karena itu, nilai F1-Score kelas True DC resolusi 1 dianggap tidak valid.

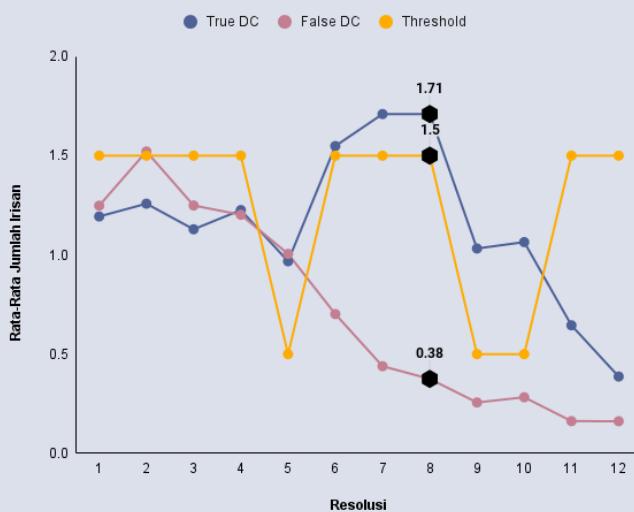
Dari gambar di atas, terlihat juga bahwa resolusi 6 memiliki waktu *running* yang cukup singkat dibandingkan dengan resolusi lainnya (resolusi 7-12) yaitu hanya selama 0,07 menit atau sekitar 4 detik untuk 1 msisdn/nomor. Selain itu, terlihat bahwa semakin tinggi resolusi H3, waktu *running* juga semakin lama. Hal tersebut disebabkan oleh semakin banyaknya *cell heksagon* hasil konversi dari *staypoint* seiring dengan semakin tingginya resolusi H3 sehingga proses perbandingan untuk mencari heksagon yang beririsan antar dua nomor juga semakin kompleks dan membutuhkan waktu yang tidak sebentar.

Berdasarkan hasil dan penjelasan di atas, dapat disimpulkan bahwa resolusi 6 adalah resolusi terbaik pada metode H3 tanpa *staypoint* karena metode ini cukup akurat dalam mengidentifikasi entitas *double counting* serta memiliki waktu *running* yang lebih singkat daripada dengan resolusi lainnya, terutama resolusi 7-12.

C. Metode *Hexagonal Hierarchical Geospatial Indexing System* (H3) dengan *Staypoint*

Berikut ini merupakan perbandingan rata-rata jumlah irisan heksagon antar kelas True DC dan kelas False DC dengan menggunakan metode *hexagonal hierarchical geospatial indexing system* (H3) dengan *staypoint* beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting



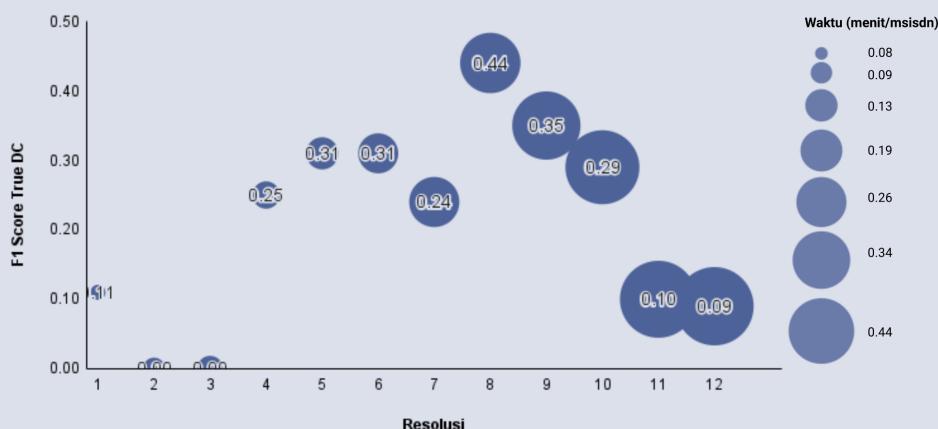
Gambar 4.76 Threshold dan Perbandingan Rata-rata Jumlah Irisan Heksagon Antar Kelas pada Data Training Algoritma Double Counting

Berdasarkan Gambar 4.76, terlihat bahwa algoritma *double counting* dengan metode H3 dengan *staypoint* dapat menghasilkan rata-rata jumlah irisan heksagon yang cukup berbeda antara kelas True DC dan False DC pada *data training*, terutama dari resolusi 6 hingga resolusi 12. Namun, diantara seluruh resolusi, resolusi 8 memiliki selisih rata-rata

jumlah irisan heksagon terbesar antara kelas True DC dan False DC. Hal ini menunjukkan bahwa metode H3 dengan *staypoint* dengan resolusi 8 dapat dengan jelas memisahkan antara kelas True DC dan False DC daripada resolusi lainnya. Mulai dari resolusi 6 hingga resolusi 12, rata-rata jumlah irisan heksagon untuk kelas True DC lebih besar dibandingkan dengan kelas False DC. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *double counting* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jumlah heksagon yang beririsan per hari lebih banyak dibandingkan dengan bukan entitas *double counting*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode H3 dengan *staypoint* dapat secara jelas memisahkan antara entitas *double counting* dan bukan entitas *double counting* pada resolusi 6 hingga resolusi 12.

Kemudian apabila dilihat berdasarkan *threshold* yang telah ditetapkan menggunakan *gini impurity*, dapat diketahui bahwa *threshold* dapat secara jelas memisahkan antara entitas *double counting* dan bukan entitas *double counting* pada resolusi 6 hingga resolusi 10. Hal tersebut ditandai dengan nilai *threshold* yang berada diantara rata-rata jumlah irisan heksagon kedua kelas. Sementara itu, *threshold* pada resolusi 1, 2, 3, 4, 5, 11, dan 12 kurang dapat memisahkan entitas *double counting* dan bukan entitas *double counting* secara jelas. Hal tersebut terlihat dari *threshold* yang bernilai lebih dari atau kurang dari rata-rata jumlah irisan heksagon kedua kelas.

Hasil Implementasi Metode H3 dengan *Staypoint* pada Data Testing Algoritma Double Counting



Gambar 4.77 Hasil Implementasi Metode H3 dengan *Staypoint* pada Data Testing Algoritma Double Counting

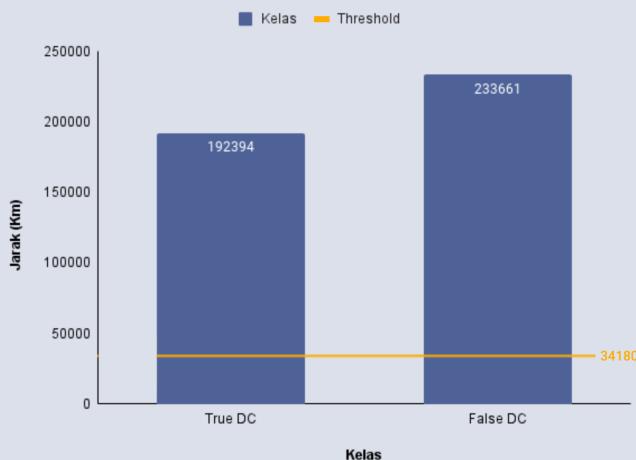
Hasil implementasi *threshold* yang telah ditetapkan terhadap *data testing* dapat dilihat pada Gambar 4.77. Berdasarkan gambar 4.77, dapat diketahui bahwa metode H3 dengan *staypoint* memiliki F1-Score kelas True DC tertinggi pada resolusi 8 yaitu sebesar 0,44. Selain itu, resolusi 8 juga memiliki waktu *running* yang cukup singkat daripada resolusi lainnya (resolusi 9-12), yaitu hanya selama 0,26 menit atau sekitar 16 detik untuk 1 msisdn/nomor. Dari grafik terlihat juga bahwa semakin tinggi resolusi H3, waktu *running* juga semakin lama. Hal tersebut disebabkan oleh semakin banyaknya *cell heksagon* hasil konversi dari *staypoint* seiring dengan semakin tingginya resolusi H3 sehingga proses perbandingan untuk mencari heksagon yang beririsan antar dua nomor juga semakin kompleks dan membutuhkan waktu yang tidak sebentar.

Berdasarkan hasil dan penjelasan di atas, dapat disimpulkan bahwa resolusi 8 adalah resolusi terbaik pada metode H3 dengan *staypoint* karena cukup akurat dalam mengidentifikasi entitas *double counting* dan memiliki waktu *running* yang lebih singkat daripada resolusi lainnya, terutama resolusi 9-12.

D. Metode *Dynamic Time Warping* (DTW)

Berikut ini merupakan perbandingan rata-rata jarak DTW antar kelas True DC dan kelas False DC beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-Rata Jarak DTW antar Kelas pada Data Training Algoritma Double Counting



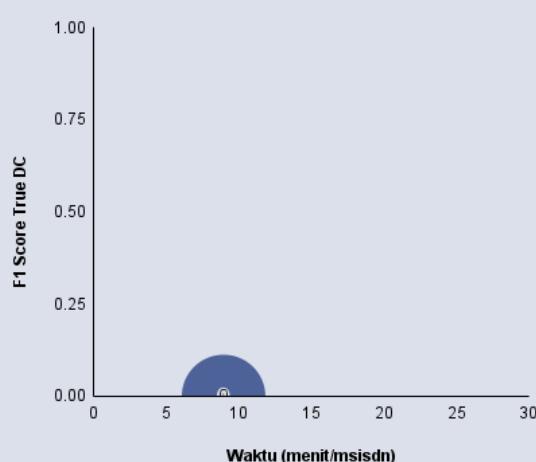
Gambar 4.78 Threshold dan Perbandingan Rata-rata Jarak DTW Antar Kelas pada Data Training Algoritma Double Counting

Berdasarkan Gambar 4.78, terlihat bahwa algoritma *double counting* dengan metode DTW dapat menghasilkan rata-rata jarak DTW yang cukup berbeda antara kelas True DC dan False DC pada *data training*. Rata-rata jarak DTW untuk kelas True DC lebih kecil dibandingkan dengan kelas False DC. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *double counting* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jarak DTW yang lebih kecil dibandingkan dengan bukan entitas *double counting*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode DTW dapat secara jelas memisahkan entitas *double counting* dan bukan entitas *double counting*.

Apabila dilihat berdasarkan *threshold* yang telah dibangun menggunakan *gini impurity*, dapat diketahui bahwa *threshold* kurang dapat memisahkan entitas *double counting* dan bukan entitas *double counting*.

Hal tersebut ditandai dengan *threshold* yang memiliki nilai kurang dari rata-rata jarak DTW kelas True DC dan kelas False DC (bukan berada diantaranya) sehingga *threshold* tersebut kurang baik dalam memisahkan entitas *double counting* dan bukan entitas *double counting* secara jelas.

Hasil Implementasi Metode DTW pada *Data Testing* Algoritma *Double Counting*



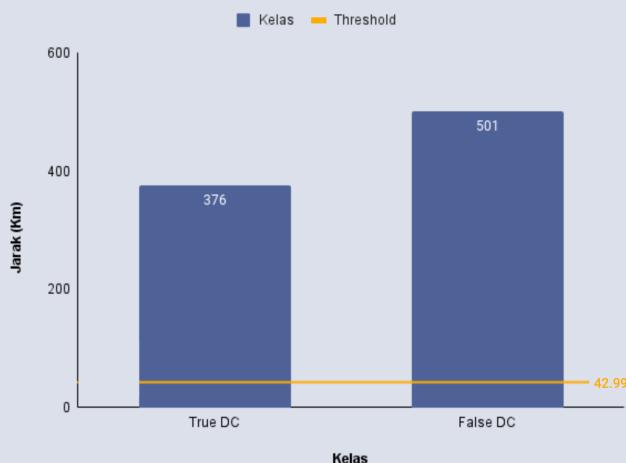
Gambar 4.79 Hasil Implementasi Metode DTW pada *Data Testing* Algoritma *Double Counting*

Hasil implementasi *threshold* yang telah ditetapkan terhadap *data testing* dapat dilihat pada Gambar 4.79. Berdasarkan gambar 4.79, dapat diketahui bahwa metode DTW memiliki F1-Score kelas True DC sebesar 0,00 dengan rata-rata waktu *running* sekitar 9 menit untuk 1 msisdn/nomor. Penyebab dari F1-Score kelas True bernilai 0,00 adalah karena *threshold* yang ditetapkan dari *data training* berada jauh dibawah rata-rata jarak DTW kelas True DC dan kelas False DC sehingga tidak mampu untuk memisahkan entitas *double counting* dan bukan entitas *double counting* pada *data testing* secara jelas. Dengan begitu, dapat disimpulkan bahwa metode DTW ini sangat tidak disarankan untuk digunakan dalam algoritma *double counting* karena memiliki nilai F1-Score kelas True DC yang sangat rendah (sangat tidak akurat) dan waktu running yang sangat lama.

E. Metode *Frechet Distance*

Berikut ini merupakan perbandingan rata-rata jarak Frechet antar kelas True DC dan kelas False DC beserta *threshold* yang telah ditetapkan.

Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Double Counting



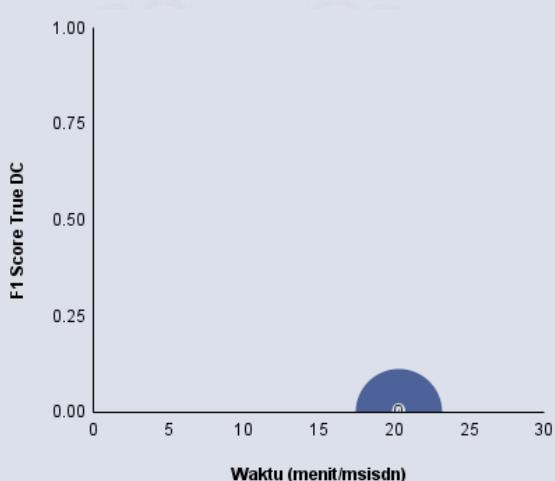
Gambar 4.80 Threshold dan Perbandingan Rata-rata Jarak Frechet Antar Kelas pada Data Training Algoritma Double Counting

Berdasarkan Gambar 4.80, terlihat bahwa algoritma *double counting* dengan metode *Frechet Distance* dapat menghasilkan rata-rata jarak Frechet yang cukup berbeda antara kelas True DC dan False DC pada *data training*. Rata-rata jarak Frechet untuk kelas True DC lebih kecil dibandingkan dengan kelas False DC. Hal ini cukup masuk akal karena memang sudah seharusnya entitas-entitas *double counting* memiliki kemiripan pola perjalanan satu sama lain sehingga menghasilkan rata-rata jarak Frechet yang lebih kecil dibandingkan dengan bukan entitas *double counting*. Hal tersebut sekaligus merupakan sebuah indikasi baik bahwa metode Frechet Distance dapat secara jelas memisahkan entitas *double counting* dan bukan entitas *double counting*.

Apabila dilihat berdasarkan *threshold* yang telah ditetapkan menggunakan *gini impurity*, dapat diketahui bahwa *threshold* kurang dapat memisahkan entitas *double counting* dan bukan entitas *double counting*.

Hal tersebut ditandai dengan *threshold* yang memiliki nilai kurang dari rata-rata jarak Frechet kelas True DC dan kelas False DC (bukan berada diantaranya) sehingga *threshold* tersebut kurang baik dalam memisahkan entitas *double counting* dan bukan entitas *double counting* secara jelas.

**Hasil Implementasi Metode *Frechet Distance* pada *Data Testing* Algoritma
*Double Counting***



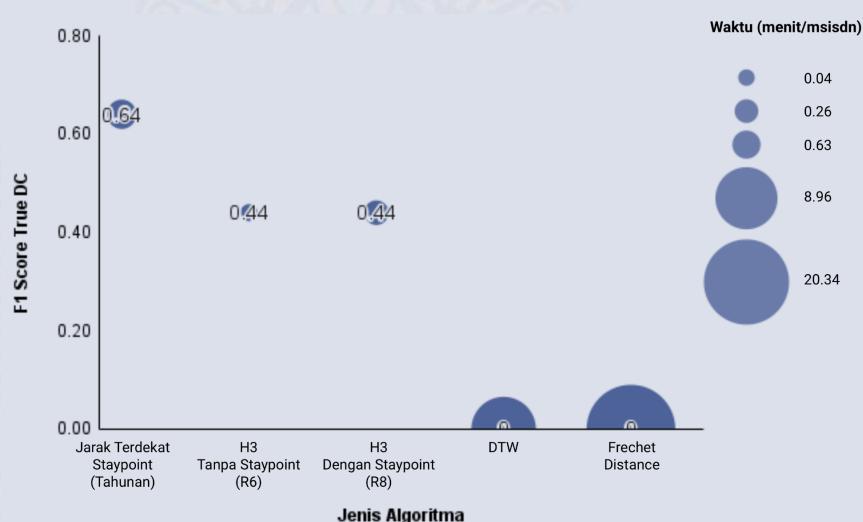
Gambar 4.81 Hasil Implementasi Metode *Frechet Distance* pada *Data Testing* Algoritma *Double Counting*

Hasil implementasi *threshold* yang telah ditetapkan terhadap *data testing* dapat dilihat pada Gambar 4.81. Berdasarkan gambar 4.81, dapat diketahui bahwa metode *Frechet Distance* memiliki F1-Score kelas True DC sebesar 0,00 dengan rata-rata waktu *running* sekitar 21 menit untuk 1 msisdn/nomor. Penyebab dari F1-Score kelas True bernilai 0,00 adalah karena *threshold* yang ditetapkan dari *data training* berada jauh dibawah rata-rata jarak Frechet kelas True DC dan kelas False DC sehingga tidak mampu untuk memisahkan entitas *double counting* dan bukan entitas *double counting* pada *data testing* secara jelas. Dengan begitu, dapat disimpulkan bahwa metode Frechet Distance ini sangat tidak disarankan untuk digunakan dalam algoritma *double counting* karena memiliki nilai F1-Score kelas True DC yang sangat rendah (sangat tidak akurat) dan waktu *running* yang sangat lama.

4.5.2.3 Perbandingan Performa Algoritma Terbaik

Algoritma terbaik dari masing-masing metode kemudian dibandingkan satu sama lain untuk menentukan algoritma yang paling baik dalam mengklasifikasikan entitas *double counting* sesuai dengan kelas aktualnya pada *data testing*. Hasil perbandingan dapat terlihat pada gambar di bawah ini.

Perbandingan Algoritma Double Counting Terbaik dari Masing-Masing Metode



Gambar 4.82 Perbandingan Algoritma Double Counting Terbaik dari Masing-masing Metode

Berdasarkan Gambar 4.82, terlihat bahwa algoritma terbaik yang dapat digunakan untuk mengatasi permasalahan *double counting* adalah algoritma kombinasi jarak terdekat antar *staypoint* dengan periode waktu referensi tahunan. Algoritma ini memiliki nilai F1-Score kelas True DC yang paling tinggi diantara algoritma lainnya. Selain itu, waktu *running* dari algoritma ini juga cukup singkat yaitu hanya selama 0,63 menit atau sekitar 38 detik untuk 1 msisdn/nomor. Walaupun terdapat algoritma lainnya yang memiliki waktu *running* lebih singkat, yaitu 0,04 menit dengan algoritma H3 tanpa *staypoint* resolusi 6 dan 0,26 menit dengan algoritma H3 dengan *staypoint* resolusi 8, tetapi nilai F1-Score kelas True DC kedua algoritma tersebut lebih kecil daripada algorirma kombinasi

jarak terdekat antar *staypoint* dengan periode waktu referensi tahunan. Sementara itu, algoritma *double counting* dengan pendekatan DTW dan *Frechet Distance* memiliki performa yang buruk dalam mengidentifikasi entitas *double counting* karena memiliki nilai F1-Score kelas True DC yang sangat rendah dan membutuhkan waktu *running* yang sangat lama. Berikut adalah *classification report* dari algoritma kombinasi jarak terdekat antar *staypoint* dengan periode waktu referensi tahunan.

Tabel 4.36 Classification Report Algoritma Kombinasi Jarak Terdekat antar Staypoint (Tahunan).

Kelas	Precision	Recall	F1-Score	Support
True DC	0,60	0,69	0,64	13
False DC	0,99	0,99	0,99	406
Macro Avg	0,80	0,84	0,82	419
Accuracy			0,98	

Berdasarkan Tabel 4.36, terlihat bahwa algoritma kombinasi jarak terdekat antar *staypoint* dengan periode waktu referensi tahunan memiliki nilai *macro average* F1-Score yang cukup tinggi yaitu sebesar 0,82. Secara keseluruhan, algoritma ini memiliki akurasi yang sangat tinggi yaitu sebesar 0,98 yang artinya algoritma ini dapat dengan tepat mengklasifikasikan 98% entitas *double counting* dan bukan entitas *double counting* sesuai dengan kelas aktualnya pada *data testing*.

4.6 Statistik Wisatawan Nusantara Asal Kota Malang

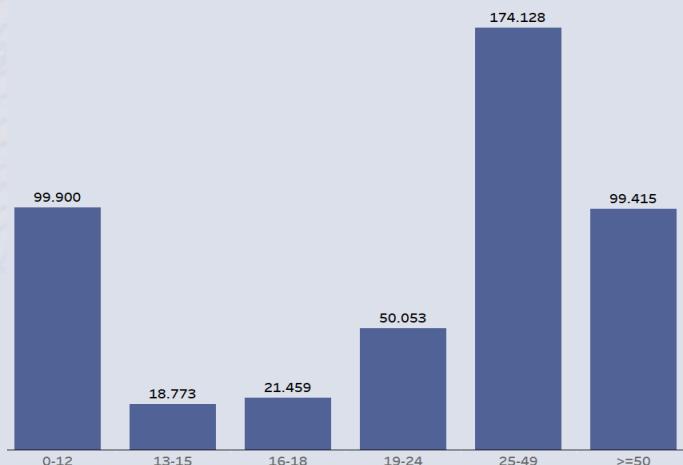
Seperti yang telah dijelaskan sebelumnya, terdapat beberapa karakteristik wisatawan yang tidak dapat dijelaskan oleh data MPD. Maka dari itu, penelitian yang dilakukan oleh Riset 1 juga meneliti tentang statistik wisatawan nusantara asal Kota Malang menggunakan survei konvensional. Penelitian ini dilakukan untuk mengetahui karakteristik wisatawan asal Kota Malang, yakni karakteristik kepemilikan *handphone*

pribadi dan kartu sim aktif, jumlah wisatawan, serta jumlah perjalanan wisata. Tidak seperti pada penelitian MPD yang telah dibahas pada subbab-subbab sebelumnya, *time reference* yang digunakan dalam penelitian statistik wisatawan nusantara ini adalah triwulan IV tahun 2022, yakni bulan Oktober—Desember tahun 2022.

4.6.1 Jumlah Wisatawan Nusantara Asal Kota Malang berdasarkan Kelompok Umur

Karakteristik wisatawan berdasarkan kelompok umur menarik untuk diperhatikan untuk melihat kecenderungan masing-masing kelompok umur dalam berwisata. Sebelum meneliti lebih jauh mengenai karakteristik wisatawan nusantara asal Kota Malang, terlebih dahulu dilakukan pengidentifikasi jumlah wisatawan nusantara berdasarkan kelompok umur dan hasilnya terlihat pada gambar 4.83.

Jumlah Wisatawan Nusantara Asal Kota Malang yang Bepergian ke Luar Kota Berdasarkan Kelompok Umur Triwulan IV 2022



Gambar 4.83 Jumlah Wisatawan Nusantara Asal Kota Malang Berdasarkan Kelompok Umur

Selama bulan Oktober—Desember 2022, secara keseluruhan terdapat sekitar 463.728 wisatawan nusantara asal Kota Malang yang bepergian ke luar kota. Kelompok umur yang mendominasi jumlah



wisatawan nusantara asal Kota Malang adalah kelompok umur 25—49 tahun yaitu sebesar 174.128 orang. Tingginya angka tersebut mengindikasikan bahwa kelompok umur 25—49 tahun memiliki antusiasme wisata yang tinggi. Sedangkan untuk kelompok umur 0—12 tahun dan di atas 50 tahun menduduki posisi kedua dan ketiga dengan jumlah wisatawan nusantara terbanyak, yaitu sebanyak 99.900 wisatawan untuk kelompok umur 0—12 tahun dan sekitar 99.415 wisatawan untuk kelompok umur lebih dari 50 tahun. Hal tersebut didukung oleh hasil salah satu studi yang dilakukan oleh Singapore Tourism Board yang menyatakan bahwa wisatawan usia lebih tua (di atas 25 tahun) lebih suka mencari pengalaman baru, unik, otentik, dan personal. Mereka sangat percaya pada ulasan-ulasan wisata terutama pada media sosial. Adapun kelompok umur 13—15 tahun adalah kelompok umur dengan jumlah wisatawan nusantara paling sedikit yaitu sekitar 18.773 orang. Dilihat dari peringkatnya, kelompok umur tersebut menjadi kelompok umur dengan antusiasme wisata paling rendah.

4.6.2 Jumlah Wisatawan Nusantara Asal Kota Malang berdasarkan Status Kepemilikan *Handphone*

Handphone menjadi alat komunikasi yang penting di era ini. Penggunaan *handphone* sangat membantu wisatawan dalam melakukan perjalanan wisata. Dengan adanya *handphone*, wisatawan menjadi lebih mudah memperoleh informasi maupun akses terhadap kegiatan pariwisata yang dilakukan serta dapat mempengaruhi preferensi perjalanan wisata wisatawan. Hal tersebut juga mendukung hasil dari survei yang dilakukan pada penelitian ini. Seperti yang terlihat pada gambar 4.84 bahwa wisatawan yang memiliki *handphone* mayoritas berada pada kelompok usia 25—49 tahun.

**Jumlah Wisatawan Nusantara Asal Kota Malang yang Bepergian ke Luar
Kota Berdasarkan Status Kepemilikan *Handphone* Menurut Kelompok
Umur Triwulan IV 2022**



Gambar 4.84 Jumlah Wisatawan Nusantara Asal Kota Malang Berdasarkan Status Kepemilikan *Handphone* dan Kelompok Umur

Pada penelitian MPD, kepemilikan *handphone* menjadi salah satu hal yang perlu diperhatikan karena terdapat kemungkinan bahwa perjalanan wisatawan yang tidak memiliki *handphone* tidak terekam dalam data MPD. Grafik di atas menunjukkan jumlah wisatawan Nusantara asal Kota Malang yang bepergian ke luar kota berdasarkan status kepemilikan *handphone* dan dikategorikan menurut kelompok umur. Berdasarkan grafik di atas, terdapat sebanyak 357.343 total wisatawan Nusantara asal Kota Malang yang memiliki *handphone* pada triwulan IV 2022. Jumlah wisatawan Nusantara yang memiliki *handphone* terbanyak berada pada kelompok umur 25—49 tahun dengan jumlah sebesar 148.796 wisatawan atau sekitar 41.64% dari seluruh kelompok umur wisatawan Nusantara yang memiliki *handphone*. Sebaliknya, hanya terdapat sekitar 4,38% wisatawan Nusantara pada kelompok umur 13—15 tahun yang memiliki *handphone* di antara seluruh wisatawan yang memiliki *handphone* pada semua kelompok umur.



Sebaran jumlah relawan atas kepemilikan ponsel pribadi didominasi oleh usia 19 tahun ke atas. Hal ini sesuai dengan hasil survei penggunaan TIK oleh Badan Litbang SDM Kominfo tahun 2017 yang mengatakan bahwa penggunaan *smartphone* paling banyak dari segi kelompok usia berada pada rentang usia 20—29 tahun, lalu diikuti oleh kelompok rentang usia 30—49 tahun. Selain itu pada rentang usia dibawah 19 tahun menunjukkan jumlah relawan yang paling sedikit atas kepemilikan *handphone* pribadi, hal ini sesuai dengan data Proporsi Individu yang Menguasai/Memiliki Telepon Genggam Menurut Kelompok Umur (Persen) yang dihitung oleh Badan Pusat Statistik pada tahun 2020 yang menyatakan bahwa persentase responden umur 15 tahun kebawah yang memiliki telepon genggam lebih sedikit dibandingkan dengan kelompok umur lainnya, yaitu hanya sebesar 24,96% saja.

Sementara itu, wisatawan Nusantara asal Kota Malang yang tidak memiliki *handphone* pada triwulan IV 2022 adalah sebanyak 106.384 wisatawan. Kelompok umur 0—12 tahun menjadi kategori kelompok umur dengan jumlah wisatawan yang tidak memiliki *handphone* terbanyak, yaitu sebanyak 45.487 wisatawan atau sekitar 42,76% dari seluruh kelompok umur wisatawan nusantara yang tidak memiliki *handphone*. Hal ini cukup rasional mengingat pada kategori kelompok umur ini merupakan kelompok umur yang terdiri dari anak-anak, balita, atau bayi yang penggunaan *handphone*-nya masih terbatas. Sebaliknya, Kelompok umur 16—18 tahun menjadi kelompok umur dengan jumlah wisatawan Nusantara yang tidak memiliki *handphone* paling sedikit, yaitu sekitar 2,33% dari keseluruhan kelompok umur yang tidak memiliki *handphone*.

Apabila ditinjau dari kepemilikan *handphone* pada masing-masing kelompok umur, terdapat sekitar 54,47% wisatawan Nusantara yang memiliki *handphone* pada kelompok umur 0—12 tahun; 83,45% pada kelompok umur 13—15 tahun; 88,47% pada kelompok umur 16—18

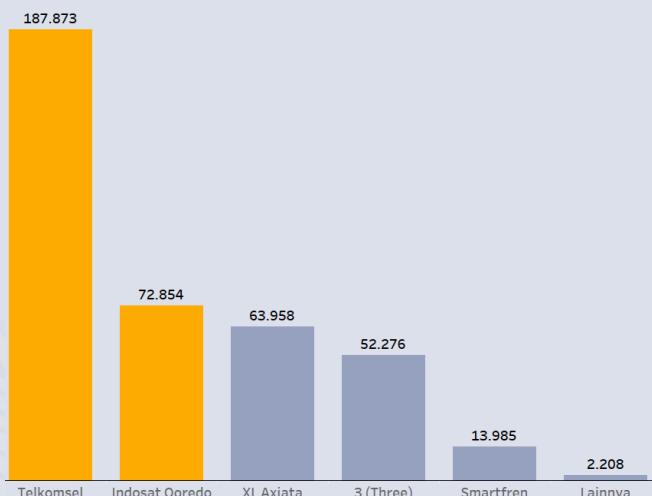
tahun; 89,16% pada kelompok umur 19—24 tahun; 85,44% pada kelompok umur 25—49 tahun; dan 75,30% pada kelompok umur 50 tahun ke atas. Hal tersebut menunjukkan bahwa persentase wisatawan Nusantara yang memiliki *handphone* jauh lebih besar daripada yang tidak memiliki *handphone* pada semua kelompok umur.

Dari hasil di atas, dapat disimpulkan bahwa jumlah wisatawan Nusantara asal Kota Malang yang bepergian ke luar kota didominasi oleh wisatawan yang memiliki *handphone*. Namun, tidak sedikit juga jumlah wisatawan Nusantara asal Kota Malang yang tidak memiliki *handphone*. Hal tersebut cukup disayangkan karena perjalanan wisata yang dilakukan wisatawan Nusantara tersebut tidak dapat direkam oleh MPD sehingga estimasi perjalanan wisata menggunakan data MPD akan menghasilkan bias.

4.6.3 Jumlah Kartu Sim Aktif Wisatawan Nusantara berdasarkan Provider

Data MPD yang berisikan catatan transaksi pengguna seluler hanya dapat diperoleh dari *provider* kartu sim yang digunakan oleh pengguna seluler, seperti Telkomsel, Indosat, XL Axiata, 3 (Three), Smartfren, dan sebagainya. Dalam upaya untuk memperbaiki kualitas dan *output* terkait perjalanan wisata yang diestimasi menggunakan data MPD, penting untuk mengetahui jumlah kepemilikan kartu sim aktif yang digunakan oleh wisatawan Nusantara berdasarkan kategori *provider* yang dipakai. Dengan diketahuinya jumlah masing-masing *provider* yang digunakan wisatawan Nusantara, untuk mendapatkan data MPD, para peneliti atau lembaga yang berfokus di bidang statistik dapat lebih memprioritaskan untuk menjalin kerja sama dengan *provider* yang lebih banyak digunakan oleh wisatawan Nusantara. Gambar 4.85 menyajikan jumlah kepemilikan kartu sim aktif yang digunakan oleh wisatawan Nusantara asal Kota Malang yang disajikan berdasarkan jenis *provider*.

**Jumlah Kepemilikan Kartu sim Aktif yang Digunakan Wisatawan
Nusantara Asal Kota Malang yang Bepergian ke Luar Kota Berdasarkan
Provider Triwulan IV 2022**



**Gambar 4.85 Jumlah Kepemilikan Kartu sim Aktif yang Digunakan Wisatawan
Nusantara Asal Kota Malang Berdasarkan Provider**

Grafik diatas menunjukkan jumlah kepemilikan kartu sim aktif yang digunakan wisatawan Nusantara asal Kota Malang yang bepergian ke luar kota berdasarkan *provider*. Telkomsel menjadi *provider* dengan jumlah kepemilikan kartu sim aktif yang paling mendominasi, yakni sebanyak 187.873 kartu sim aktif. Kemudian Indosat menempati posisi kedua dengan jumlah kepemilikan kartu sim aktif sebanyak 72.854 kartu sim aktif. Adapun *provider* dengan jumlah kepemilikan yang paling sedikit adalah kategori *provider* lainnya (*provider* selain Telkomsel, Indosat, XL Axiata, 3 (Three), dan Smartfren), yakni hanya sekitar 2.208.

Perbandingan jumlah kepemilikan kartu sim aktif menunjukkan adanya perbedaan yang cukup signifikan antar *provider*. Terlihat bahwa Telkomsel dan Indosat memiliki posisi lebih unggul dibandingkan dengan *provider* lainnya. Bahkan, jumlah kepemilikan kartu sim aktif Telkomsel hampir mencapai tiga kali lipat dari jumlah kepemilikan kartu sim aktif XL Axiata. Indosat berada di posisi kedua dengan jumlah kartu sim aktif yang juga signifikan. Hal ini menunjukkan bahwa kedua *provider* tersebut

cukup populer di kalangan wisatawan Nusantara asal Kota Malang. Sehingga para peneliti atau lembaga yang berfokus di bidang statistik dapat lebih memprioritaskan untuk menjalin kerja sama dengan *provider* Telkomsel dan Indosat untuk memperoleh data MPD wisatawan nusantara. Selain itu, kerja sama dengan *provider* XL Axiata juga dapat menjadi salah hal yang dipertimbangkan, mengingat berdasarkan data di atas, XL Axiata termasuk ke dalam 3 besar *provider* dengan jumlah kepemilikan kartu sim terbanyak.

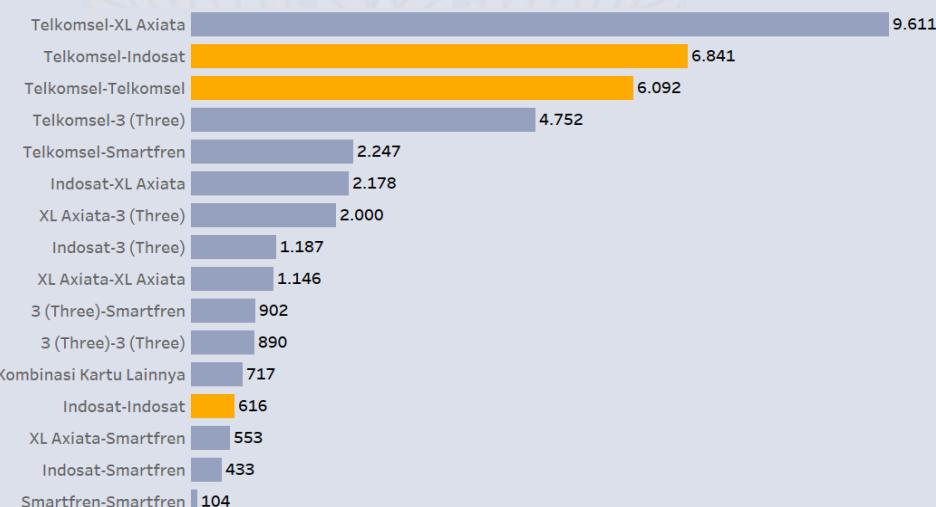
Banyaknya jumlah relawan yang menggunakan kartu sim Telkomsel dan Indosat secara aktif sejalan dengan *Data Indonesia Bets on 4G to Increase Mobile Internet Coverage; 5G to Focus on Industry* Tahun 2022 oleh Counterpointsearch. Dalam artikel tersebut, dinyatakan bahwa *provider* Telkomsel dan Indosat secara berurutan menguasai pasar MNO Q1 2022 di Indonesia. Hasil yang sama juga didapatkan oleh survei *Mobile Phone Usage and 5G Network Projection* di Indonesia yang menyebutkan bahwa Telkomsel menjadi *provider* favorit masyarakat Indonesia. Selain itu, *provider* Telkomsel juga sudah menyediakan layanan jaringan 5G yang banyak diminati masyarakat (Dhini, 2022). Dalam survei *Mobile Phone Usage and 5G Network Projection* ini, didapatkan bahwa persentase responden yang memilih Telkomsel mencapai 46%. Lalu, diikuti Indosat Hutchison 33%, XL Axiata 17%, dan Smartfren 4%.

4.6.4 Jumlah Wisatawan Nusantara dengan Kepemilikan Ganda Kartu sim Aktif dalam Handphone Utama

Seperti yang sudah dijelaskan sebelumnya, terdapat kemungkinan terjadinya *double counting* dalam data *Mobile Positioning Data* (MPD). *Double counting* harus dihindari karena hal tersebut akan berpengaruh terhadap jumlah wisatawan yang menggunakan dua kartu sim aktif dalam *handphone*-nya ketika melakukan perjalanan wisata. Dengan adanya *double counting*, satu wisatawan dapat terhitung menjadi dua wisatawan dan tentu saja hal tersebut dapat menyebabkan *overestimate* jumlah

wisatawan yang diestimasi menggunakan data MPD. Untuk mendukung pemecahan masalah *double counting* ini, diperlukan suatu gambaran mengenai jumlah wisatawan Nusantara yang menggunakan dua kartu sim aktif di dalam satu *handphone* utama berdasarkan kategori kombinasi *provider* yang dipakai. Penelitian ini juga mengumpulkan data mengenai hal tersebut dan hasilnya terlihat pada gambar di bawah ini.

**Jumlah Wisatawan Nusantara Asal Kota Malang yang Bepergian ke Luar
Kota dengan Kepemilikan Ganda Kartu sim Aktif dalam Satu *Handphone*
Utama Berdasarkan Kombinasi *Provider* Triwulan IV 2022**



**Gambar 4.86 Jumlah Wisatawan Nusantara Asal Kota Malang dengan
Kepemilikan Ganda Kartu sim Berdasarkan Kombinasi *Provider***

Sesuai dengan *provider* utama yang dianalisis dalam penelitian tentang MPD pada penelitian ini, fokus utama kombinasi *provider* pada gambar di atas adalah kombinasi dari *provider* Telkomsel dan Indosat. Berdasarkan gambar di atas, terlihat bahwa jumlah wisatawan Nusantara asal Kota Malang yang menggunakan kombinasi *provider* Telkomsel-Indosat dan Telkomsel-Telkomsel berada pada urutan kedua dan ketiga di bawah kombinasi *provider* Telkomsel-XL Axiata. Selain itu, penggunaan dua kartu sim Indosat tampaknya tidak terlalu diminati, hal ini terlihat dari jumlah pengguna dua kartu Indosat sekaligus yang hanya

terdapat sebanyak 616 wisatawan atau menempati urutan ke-13 dari enam belas jenis pasangan *provider* yang digunakan oleh wisatawan Nusantara asal kota Malang.

Selain dari kombinasi *provider* yang menjadi fokus utama, terlihat juga bahwa kombinasi *provider* Telkomsel-XL Axiata menempati urutan pertama. Hal tersebut memiliki arti bahwa jumlah wisatawan Nusantara asal Kota Malang didominasi oleh pengguna kombinasi *provider* Telkomsel-XL Axiata dengan jumlah wisatawan sebanyak 9.611 orang. Hasil ini turut menyimpulkan bahwa untuk penelitian selanjutnya, disarankan agar menambahkan *provider* XL Axiata ke dalam penelitian.

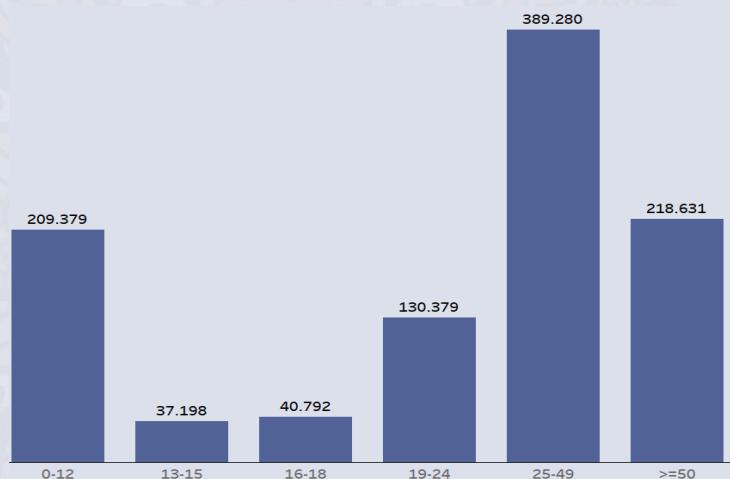
4.6.5 Jumlah Perjalanan Wisata Asal Kota Malang berdasarkan Kelompok Umur

Pada pembahasan sebelumnya, sudah dijelaskan mengenai jumlah wisatawan nusantara berdasarkan kelompok umur. Selain jumlah wisatawan, jumlah perjalanan wisata berdasarkan kelompok umur juga menarik untuk diperhatikan untuk melihat kecenderungan masing-masing kelompok umur dalam berwisata. Wisatawan adalah seorang pengunjung yang menetap di suatu tempat yang dikunjungi minimal 24 jam tetapi tidak lebih dari dua belas bulan dengan tujuan pribadi (liburan, beristirahat, mengunjungi kerabat, pendidikan, pelatihan, dll) dan bisnis (konferensi, pameran, seminar, kegiatan jual beli, dll) (Badan Pusat Statistik, 2021). Singkatnya, setiap orang yang mengunjungi suatu tempat atau berwisata, dengan beberapa syarat yang disebutkan dalam pengertian sebelumnya, dapat disebut sebagai wisatawan.

Jumlah wisatawan dapat digunakan sebagai istilah yang menggambarkan total banyaknya wisatawan. Untuk dapat dikatakan sebagai wisatawan, seseorang harus melakukan perjalanan wisata. Sebuah perjalanan wisata dapat terdiri dari beberapa kunjungan (*visits*) ke berbagai

tempat, namun jumlah perjalannya tetap dihitung satu. Satu perjalanan dihitung mulai dari seseorang berangkat dari lingkungan kesehariannya ke suatu tempat tertentu hingga orang tersebut kembali ke lingkungan kesehariannya (IRTS, 2008). Penelitian ini hanya menghitung jumlah perjalanan asal Kota Malang yang bepergian ke luar kota, yakni perjalanan yang dilakukan dari Kota Malang menuju kota/kabupaten lain. Jumlah perjalanan wisata asal Kota Malang yang bepergian ke luar kota dapat terlihat pada gambar 4.87.

Jumlah Perjalanan Wisata Asal Kota Malang yang Bepergian ke Luar Kota Berdasarkan Kelompok Umur Triwulan IV 2022



Gambar 4.87 Jumlah Perjalanan Wisata Asal Kota Malang Berdasarkan Kelompok Umur

Pada triwulan IV tahun 2022, jumlah perjalanan wisata asal kota Malang yang bepergian ke luar kota didominasi oleh kelompok umur 25—49 tahun, yakni sebanyak 389.280 perjalanan. Sementara itu, kelompok umur dengan jumlah perjalanan paling sedikit adalah penduduk berusia 13—15 tahun, yakni sebanyak 37.198 perjalanan. Selain itu, penduduk pada kelompok umur 0—12 tahun termasuk ke dalam 3 kelompok umur dengan jumlah perjalanan wisata terbanyak. Berdasarkan gambar di atas, dapat dikatakan bahwa penduduk pada kelompok umur 25—49 tahun memiliki antusiasme berwisata paling tinggi dibandingkan

dengan kelompok umur lainnya. Sebaliknya, kelompok umur dengan antusiasme berwisata paling rendah adalah penduduk dengan rentang usia 13—15 tahun.

Apabila dibandingkan dengan jumlah wisatawan Nusantara asal Kota Malang yang telah dibahas sebelumnya pada gambar 4.84, terlihat bahwa kedua grafik yang dihasilkan memiliki pola yang cukup mirip. Keduanya memiliki frekuensi paling tinggi pada kelompok umur 25—49 tahun, sedangkan frekuensi paling rendah pada kelompok umur 13—15 tahun. Namun, pada grafik jumlah wisatawan Nusantara terlihat bahwa jumlah wisatawan dalam kelompok umur 0—12 tahun lebih tinggi daripada dalam kelompok umur lebih dari sama dengan 50 tahun, sedangkan pada grafik jumlah perjalanan wisata, kelompok umur 0—12 tahun terlihat lebih rendah. Hal tersebut menunjukkan bahwa wisatawan Nusantara asal Kota Malang yang berumur lebih dari 50 tahun lebih sering melakukan perjalanan wisata ke luar kota daripada wisatawan nusantara asal Kota Malang yang berumur 0—12 tahun.

4.6.6 Jumlah Perjalanan Wisata Asal Kota Malang berdasarkan Bulan

Selain berdasarkan kelompok umur, jumlah perjalanan wisata pada masing-masing bulan selama tiga bulan terakhir tahun 2022 juga menarik untuk didalami untuk mengetahui apakah terdapat perbedaan jumlah perjalanan wisata pada bulan *high* dan *low season*. Selain itu, membandingkan jumlah perjalanan yang didapatkan dari hasil survei konvensional terhadap jumlah perjalanan yang didapatkan dari MPD juga menarik dilakukan. Hasil eksplorasi kedua hal tersebut tersaji dalam gambar 4.88.

Jumlah Perjalanan Wisata Bulanan Asal Kota Malang yang Bepergian ke
Luar Kota Triwulan IV 2022



Gambar 4.88 Jumlah Perjalanan Wisata Bulanan Asal Kota Malang

Berdasarkan grafik di atas, terlihat bahwa terdapat perbedaan jumlah perjalanan wisata asal Kota Malang yang bepergian ke luar kota untuk masing-masing bulan selama triwulan IV 2022. Menurut survei konvensional, jumlah perjalanan tertinggi ada pada bulan Oktober yang mencapai angka 497.328 perjalanan dan jumlah perjalanan terendah ada pada bulan November yang hanya mencapai angka 60.817 perjalanan.

Jika dibandingkan dengan jumlah perjalanan yang bersumber dari data MPD, terdapat perbedaan yang signifikan pada data yang dihasilkan oleh kedua sumber data. Jumlah perjalanan yang terekam dari data MPD di bulan November adalah sebanyak 516.147 perjalanan, sementara data yang berhasil ditangkap dalam survei konvensional hanya mencapai puluhan ribu saja. Tidak hanya itu, data MPD mencatat bahwa jumlah perjalanan wisata bulanan pada bulan Desember adalah yang tertinggi, yaitu 618.740 perjalanan, berbeda dengan data survei konvensional yang hasil estimasinya menyatakan terdapat 467.514 perjalanan dan bukan menjadi bulan dengan jumlah perjalanan wisata tertinggi. Kesamaan pola perjalanan antara kedua sumber data hanya terlihat pada data bulan

November yang sama-sama tercatat sebagai perjalanan paling sedikit sepanjang triwulan terakhir 2022.

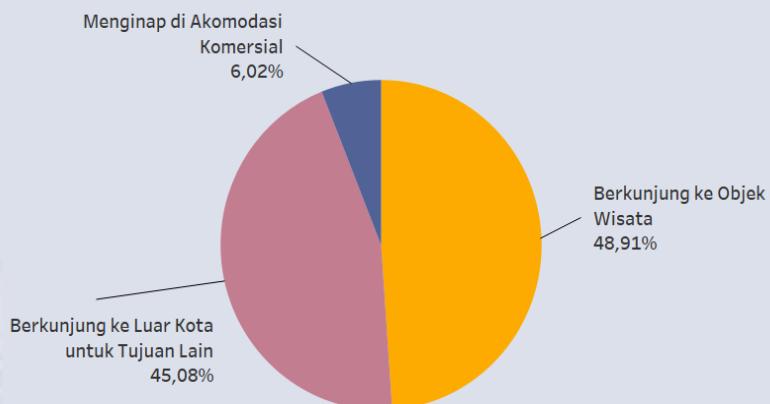
Pola perjalanan data MPD sejalan dengan pola perjalanan wisata bulanan yang dihasilkan oleh BPS pada tahun 2021, yakni jumlah perjalanan triwulan terakhir berturut-turut adalah 62.821.537 perjalanan di bulan Oktober, 60.989.058 perjalanan di bulan November, dan 63.206.792 perjalanan di bulan Desember (Badan Pusat Statistik, 2021). Perbedaan data jumlah perjalanan ini cukup menarik untuk ditinjau lebih lanjut mengingat bahwa metode yang dipakai dari kedua sumber cukup berbeda. Suatu penelitian mengatakan bahwa jumlah hari libur berhubungan positif terhadap permintaan berwisata (Mustika, 2019). Data MPD cenderung berfluktuasi mengikuti jumlah hari libur bulanan dibandingkan data survei. Karena itu, data MPD dapat mengatasi kelemahan yang muncul dari survei konvensional. Kelemahan tersebut diantaranya berupa adanya keterbatasan daya ingat responden dan *non-sampling error*. *Non-sampling error* adalah kejadian dimana terdapat bias dan tidak akurat pada jawaban responden (Lohr, 1897). Survei Wisatawan Nusantara yang ingin mengumpulkan keterangan dengan level mikro lebih disarankan untuk memanfaatkan periode ingatan yang lebih pendek karena *recall error* responden semakin meningkat seiring dengan semakin lamanya periode peristiwa yang ingin diingat (Kjellsson et al., 2014). Selain itu, data MPD dapat menangkap riwayat perjalanan secara lebih banyak dibandingkan dengan menggunakan metode konvensional sehingga untuk pengumpulan data perjalanan ke depannya, penggunaan data MPD dapat dipertimbangkan.

4.6.7 Persentase Perjalanan Wisata Asal Kota Malang berdasarkan Kategori Perjalanan Wisata

Setelah mengetahui jumlah perjalanan wisata asal Kota Malang yang bepergian ke luar kota pada masing-masing bulan di triwulan IV 2022, muncul pertanyaan lanjutan, yakni kemana sajakah perjalanan tersebut dilakukan. Untuk menjawab pertanyaan tersebut, peneliti juga

menghitung persentase perjalanan wisata asal Kota Malang berdasarkan kategori perjalannya, seperti yang terlihat pada gambar 4.89.

**Persentase Perjalanan Wisata Asal Kota Malang yang Bepergian ke Luar
Kota Berdasarkan Kategori Perjalanan Wisata Triwulan IV 2022**



Gambar 4.89 Perjalanan Wisata Asal Kota Malang Berdasarkan Kategori Perjalanan Wisata

Berdasarkan grafik di atas, dapat diketahui bahwa sebanyak 48,91% perjalanan wisata asal Kota Malang yang bepergian ke luar kota dikategorikan sebagai perjalanan wisata karena responden mengunjungi objek wisata. Kemudian, sebanyak 6,02% perjalanan wisata asal Kota Malang dikategorikan sebagai perjalanan wisata karena responden menginap di akomodasi komersial, sedangkan 45,08% sisanya dikategorikan sebagai perjalanan wisata karena responden berkunjung ke luar kota dengan tujuan selain mengunjungi objek wisata dan menginap di akomodasi komersial.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan terkait rancangan algoritma dan hasil Survei Wisatawan Nusantara yang telah dijelaskan pada Bab IV, diperoleh kesimpulan sebagai berikut.

1. Sesuai tujuan yang pertama, penelitian ini telah menerapkan penjaminan kualitas (*quality assurance*) data MPD yang dihasilkan dari nomor relawan. Prosedur yang dilakukan diantaranya adalah melakukan eksplorasi terhadap distribusi jumlah *records*, jumlah *subscribers*, dan jumlah *records duplicate*; normalisasi data atribut *latitude*, *longitude*, dan *location area coordinate* (LAC); dan eksplorasi jumlah provinsi, kabupaten, dan kecamatan unik perhari. Dari kegiatan yang telah dilakukan, dapat disimpulkan bahwa data MPD dari kedua *provider* berada pada kondisi yang cukup stabil dan lengkap. Meskipun ditemukan beberapa nilai null pada atribut *latitude*, *longitude*, dan LAC, namun data kedua *provider* tersebut masih dapat dikatakan baik dengan kategori data sangat lengkap dan lengkap sempurna.
2. Algoritma *family grouping* terbaik untuk penelitian ini adalah Hexagonal Hierarchical Geospatial Indexing System (H3geo) dengan *staypoint* dan resolusi 10. Algoritma ini merupakan algoritma dengan performa terbaik berdasarkan hasil evaluasi yang menunjukkan nilai F1-score sebesar 88% dan akurasi 98%. Hexagonal Hierarchical Geospatial Indexing System (H3) merupakan sistem pengindeksan geospasial untuk mengelompokkan data spasial berdasarkan *grid* yang berbentuk heksagonal. Algoritma ini membuat pengelompokan menjadi lebih baik karena bentuk *grid* heksagon dapat menghasilkan perhitungan yang konsisten dan membuat semua tetangga memiliki jarak yang sama. Algoritma H3 dengan *staypoint* menghasilkan



performa yang lebih baik daripada tanpa *staypoint* karena tidak semua koordinat lokasi dikonversi menjadi indeks H3, melainkan hanya koordinat lokasi yang dijadikan *staypoint*. Sebuah entitas dikatakan sebagai entitas *family grouping* jika memiliki rata-rata jumlah irisan heksagon antar nomor minimal sama dengan nilai *threshold* yang terlebih dahulu ditentukan dengan memanfaatkan *gini impurity*.

3. Algoritma *double counting* terbaik untuk penelitian ini adalah algoritma kombinasi jarak terdekat antar *staypoint* dengan skema data tahunan. Algoritma ini merupakan algoritma dengan performa terbaik berdasarkan hasil evaluasi yang menunjukkan nilai F1-score sebesar 64% dan akurasi 98%. Algoritma ini merupakan algoritma yang didasari pada pengukuran jarak antar *staypoint* pada setiap nomor *subscribers* dengan lokasi *home* yang sama menggunakan jarak *Haversine*. Jarak *Haversine* dipilih karena memperhitungkan bahwa bumi adalah bidang yang memiliki derajat kelengkungan dan bukan sebuah bidang datar. Jika beberapa nomor memiliki jarak yang relatif dekat, maka nomor-nomor tersebut diindikasikan merepresentasikan *subscribers* yang sama. Sebuah entitas dikatakan sebagai entitas *double counting* jika rata-rata kestabilan jarak antar nomor bernilai kurang dari suatu nilai *threshold* yang terlebih dahulu ditentukan dengan memanfaatkan *gini impurity*.
4. Melalui pendekatan *mobile positioning data* (MPD) dan survei konvensional wisatawan nusantara, diperoleh statistik wisatawan nusantara di Kota Malang selama Triwulan IV 2022. Jumlah wisatawan nusantara didominasi oleh kelompok umur 25 – 49 tahun yang berarti bahwa kelompok umur tersebut memiliki antusiasme wisata relatif lebih tinggi daripada kelompok umur lainnya. Wisatawan nusantara tersebut mayoritas memiliki *handphone* dengan *provider* yang paling banyak digunakan adalah Telkomsel dan Indosat. Meskipun begitu, kombinasi *provider* Telkomsel–XL Axiata merupakan kombinasi *provider* yang paling banyak digunakan oleh wisatawan. Berdasarkan hasil survei,

ditemukan bahwa mayoritas perjalanan dikategorikan sebagai perjalanan wisata karena wisatawan mengunjungi objek wisata. Jika melihat data pada masing-masing bulan, baik berdasarkan data MPD maupun data hasil survei, jumlah perjalanan wisata di bulan November merupakan yang paling rendah dibandingkan dengan bulan Oktober dan Desember. Hal tersebut dapat disebabkan oleh banyaknya hari libur di kedua bulan tersebut (Oktober dan Desember). Menariknya, terdapat perbedaan yang cukup signifikan antara jumlah perjalanan wisata yang tercatat pada data MPD dan data hasil survei. Perbedaan yang paling signifikan terjadi pada jumlah perjalanan wisata di bulan November.

5.2 Saran

Berdasarkan penelitian yang sudah dilakukan serta hasil dan pembahasan yang sudah dijabarkan, maka terdapat beberapa saran yang diajukan sebagai berikut:

1. Bagi Badan Pusat Statistik

BPS dapat menggunakan algoritma terbaik *family grouping* dan *double counting* yang telah dibangun pada penelitian ini sebagai proses tambahan dalam mengolah data sebelum dilakukan penghitungan estimasi Statistik Wisatawan Nusantara agar menghasilkan estimasi yang lebih akurat. Dalam penerapannya, penghitungan Statistik Wisatawan Nusantara akan lebih baik dilakukan dengan kombinasi antara MPD dan Survei Konvensional karena terdapat beberapa karakteristik wisatawan yang tidak bisa didapatkan hanya melalui MPD.

2. Bagi Penelitian Selanjutnya

Penelitian selanjutnya dapat melakukan riset yang serupa dengan penelitian ini, tetapi dengan melakukan penelitian dan eksplorasi lebih lanjut terhadap metode-metode lainnya yang dapat digunakan dalam penghitungan similaritas antar nomor dan juga metode yang digunakan untuk pembangunan *threshold*. Selain itu, penelitian lainnya dapat



melakukan validasi lebih lanjut terhadap hasil penelitian ini dengan menggunakan sampel data MPD yang lebih besar dan melibatkan wilayah yang lebih luas untuk memberikan kepercayaan yang lebih tinggi terhadap hasil penelitian dan memperluas generalisabilitas temuan.

DAFTAR PUSTAKA

- Ahas, R., Aasa, A., Slim, S., & Aunap, R. (2007). Mobile Positioning in Space-Time Behaviour Studies: Social Positioning Method Experiments in Estonia. *Cartography and Geographic Information Science*, 34(Computer Science), 259-273.
<https://doi.org/10.1559/152304007782382918>
- Ahas, R., Silm, S., Järv, O., Saluveer, E., & Tiru, M. (2010). Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, 17, 3-27.
- Ahas, et. al., R. (2007). *Mobile Positioning Data in Tourism Studies and Monitoring: Case Study in Tartu, Estonia*, 120.
- Ahmad Roihan, dkk, *Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper*, IJCIT, Vol.5 No.1, 2019, hlm.75-82.
- Alt, H., & Godau, M. (1995). *COMPUTING THE FRÉCHET DISTANCE BETWEEN TWO POLYGONAL CURVES* (Vol. 05(01n02), 75–91). International Journal of Computational Geometry & Application. 10.1142/s0218195995000064
- Andrey. (2022). H3 Hexagonal Grid: Why We Use It for Data Analysis and Visualization.
- Arikunto, S. (2010). Prosedur Penelitian Suatu Pendekatan Praktik. Jakarta: PT. Rineka Cipta.
- Awangga, R. M. (2019). *Dasar-Dasar Python* (Vol. 2019 hlm. 1).
- Badan Pusat Statistik. (2019). *Tabel Dinamis Proporsi Kontribusi Pariwisata Terhadap PDB 2015-2017*. Badan Pusat Statistik.
<https://www.bps.go.id/indicator/16/1188/1/proporsi-kontribusi-pariwisata-terhadap-pdb.html>
- Badan Pusat Statistik. (2023). *Tabel Dinamis Jumlah Perjalanan Wisatawan Nusantara 2018-2020*. Badan Pusat Statistik.
<https://www.bps.go.id/indicator/16/1189/1/jumlah-perjalanan-wisatawan-nusantara.html>

Badan Pusat Statistik. (2023). *Tabel Dinamis Proporsi Individu yang Menguasai/Memiliki Telepon Genggam Menurut Kelompok Umur (Persen), 2020-2022.* Badan Pusat Statistik. <https://www.bps.go.id/indicator/27/1222/1/proporsi-individu-yang-menggunakan-telepon-genggam-menurut-kelompok-umur.html>

Badan Pusat Statistik. (2019). *Statistik Objek Daya Tarik Wisata Tourist Attraction Object Statistics.*

Badan Pusat Statistik. (2021). *Statistik Wisatawan Nusantara 2020.* Badan Pusat Statistik.

Badan Pusat Statistik. (2022). *Statistik Wisatawan Nusantara 2021.* Badan Pusat Statistik.

<https://www.bps.go.id/publication/2022/09/29/63fea546c28b8eb2c3c71004/statistik-wisatawan-nusantara-2021.html>

Baker, J., & Sugden, S. (2003). Spreadsheets in Education The First 25 Years. *Spreadsheets in Education, 1(1)*, 18-43.

Bean, J. (2022). Creating Consistent Spatial Calculations with Uber's H3 Hexagons.

<https://towardsdatascience.com/creating-consistent-spatial-calculations-with-ubers-h3-hexagons-1af032802a77>

Beard, L., & Aghassibakes, N. (2021). Tableau (version 2020.3). *Journal of the Medical Library Association, 109(1)*, 159–161. <https://doi.org/10.5195/jmla.2021.1135>

Belloum, A. (n.d.). *Introduction to Big Data.* https://hpc.uva.nl/uploaded_files/inlineitem/Lecture_6_BigData.pdf

BlazStrle, Mozina, M., Bratko, I. (2003). Qualitative Approximation to Dynamic Time Warping Similarity between Time Series Data. University of Ljubljana. Slovenia.

Bojic, I., Massaro, E., Belyi, A., Sobolevsky, S., & Ratti, C. (2015, Desember). Choosing the right home location definition method for the given dataset. *International Conference on Social Informatics*, 194 - 208.

Brahma, D. (2020). Cara Menggunakan Google Colaboratory. *In Medium.* <https://medium.com/@dede.brahma2/cara-menggunakan-google-colaboratory-5f5e4393ac2f>

- Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G., Bin, De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, 6, 61677–61685. <https://doi.org/10.1109/ACCESS.2018.2874767>
- Chaplowe, S. G. (n.d.). *Counting People Reached – Navigating direct, indirect and double counting*. <https://scottchaplowe.com/counting-people-reached/>
- C. L. Philip Chen, & C. Y. Zhang. (2014). *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data*, 314-347.
- Cuatomsportal, E. (2019). *Definition: Usual Environment*. tariffnumber.com
- Dang, T. A., Deepak, J., Wang, J., Luo, S., Jin, Y., Ng, Y., Lim, A., & Li, Y. (2017). Mobility Genome - A Framework for Mobility Intelligence from Large-Scale Spatio-Temporal Data. *2017 International Conference on Data Science and Advanced Analytics*. 10.1109/DSAA.2017.20
- Derdouri, A., & Osaragi, T. (2021). *A Machine Learning-Based Approach for Classifying Tourists and Locals Using Geotagged Photos: The Case of Tokyo* (Vol. 23, 2021, hlm. 575-609). Springer.
- ETSI. (2000). *European digital cellular telecommunications system (phase 2); numbering, addressing, and identification (gsm 03.03 version 4.10.1)* (Vol. 4). France: European Telecommunications Standards Institute 2000.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Technique*. Berlin : Springer.
- Guo. (2017). *An Efficient Query Algorithm for Trajectory Similarity Based on Fréchet Distance Threshold* (Vol. 6(11), 326). SPRS International Journal of Geo-Information. 10.3390/ijgi6110326
- Hasanuddin, A. A. (2020). Rancang Bangun Web-GIS Berbasis Geodjango-Python. Doctoral dissertation, Universitas Hasanuddin.
- Han, J., Hedblom, A., & N, N. (2011). *A Density-Based Spatial Clustering of Application with Noise*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (Third Edition). United State of America: Elsevier.

Hartama, D. (2018). Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 3, 46. <https://doi.org/10.30645/jurasik.v3i0.65>

He, S. (2018). *From Beautiful Maps to Actionable Insights_ Introducing kepler.* <https://www.uber.com/en-ID/blog/keplergl/>

I Made, I., & Cahyaningrum, I. (2020). Metodologi Penelitian Pendidikan.

IRTS. (2008). *International Recommendations for Tourism Statistics*. Department of Economic and Social Affairs Statistics Division. <https://unstats.un.org/unsd/trade/IRTS/IRTS%202008%20unedited.pdf>

Khazari. (2017). *Sistem Rekomendasi Penentuan Judul Skripsi Menggunakan Algoritma Decision Tree*. Fakultas Teknik, Universitas WidyaGama.

Kjellsson, G., Clarke, P., & Gerdthama, U.-G. (2014). Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. *Journal of Health Economics*, 35, 34 - 46. <http://dx.doi.org/10.1016/j.jhealeco.2014.01.007>

Koch, P. R., Hölbig, C. A., & Rieder, R. (2018). *Using Kepler.gl to visualize weather data*.

Koranti, Komsi, Sriyanto, & Lestiyono, S. (2017). Analisis Preferensi Wisatawan Terhadap Sarana di Wisata Taman Wisata Kopeng. *Jurnal Ekonomi Bisnis*, 22. <https://core.ac.uk/display/154930852>

Kunicki, Z. J., Zambrotta, N. S., Tate, M. C., Surrusco, A. R., Risi, M. M., & Harlow, L. L. (2019). Keep Your Stats in the Cloud! Evaluating the Use of Google Sheets to Teach Quantitative Methods. *Journal of Statistics Education*, 27(3), 188-197. <https://doi.org/10.1080/10691898.2019.1665485>

Kurniawan, H. (n.d.). *Aplikasi Penjawab Pesan Singkat Automatis Dengan Bahasa Phyton* (Vol. 2005 hlm. 1). UNDIP.

Lendave, V. (2021). *Gini Impurity vs Information Gain vs Chi-Square – Methods for Decision Tree Split*. <https://analyticsindiamag.com/>

Levitin, A. (2006). Introduction to the Design and Analysis of Algorithms, 2d, ed.

Lhor, S. L. (n.d.). Sampling and Nonsampling Errors. In *m Sampling: Design and Analysis* by Sharon L. Lohr (p. 15).

https://statacumen.com/old/courses/stat345/handouts/handout_sampling.pdf

- Lutz, m. (2010). *Programming Python, 4th Edition*. O'Reilly Media, Inc.
- Matteson, S. (2013). *Big Data basic concepts and benefits explained*. www.techrepublic.com
- Mazza, R. (2009). Introduction to information visualization. Springer Science & Business Media.
- McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. <https://doi.org/10.1080/14649373.2013.831195>
- Meyer, D. Z., & Avery, L. M. (2009). Excel as a Qualitative Data Analysis Tool. *Field Methods*, 21(1), 91-112. <https://doi.org/10.1177/1525822X08323985>
- Mirajanatin, H. P., Sutikno, F. R., & Sari, N. (2013). *Kajian Potensi Pariwisata Perkotaan Di Kota Malang Berdasarkan Stakeholder*.
- Mumtaz, K., & Duraiswamy, K. (2010). *An Analysis On Density Based Clustering Of Multi Dimensional Spatial Data* (PP 8-12 ed.). Indian Journal Of Computer Science and Engineering.
- Munaf, A. R. M., Putra, A. P., Madjida, W. O. Z., Setyadi, I. A., & Nugroho, A. R. S. (2021). Data Input Quality Metrics on Mobile Positioning Data (MPD). In *Proceedings of The International Conference on Data Science and Official Statistics* (Vol. 2021, No. 1, pp. 833-841).
- Mustika, T. I. (2019). ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PENDAPATAN ASLI DAERAH SEKTOR PARIWISATA DI KABUPATEN BANTUL TAHUN 2014. <http://repository.ums.ac.id/bitstream/handle/123456789/29241/Naskah%20Publikasi.pdf?sequence=12&isAllowed=y>
- Nations, U. (2010). *International Recommendations for Tourism Statistics 2008*.
- Nations, U., & Unies, N. (2007). *International Recommendations on Tourism Statistics (IRTS) Provisional draft. September*.
- Nugroho, A. R. S., Munaf, A. R., Madjida, W. O. Z., Putra, A. P. P., & Setyadi, I. A. (2021). Home and work identification process using mobile positioning data. In *Conference of European Statisticians: Expert Meeting on Statistical Data Collection: UNECE*.



Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support Vector Machine dan Aplikasinya Dalam Bioinformatika. <http://www.ilmukomputer.com>

Octasylva, Pratiwi, A. R., & Rurianto, J. (2020). Analisis Industri Telekomunikasi Seluler di Indonesia: Pendekatan SCP (Structure Conduct Performance). *INOBIS: Jurnal Inovasi Bisnis dan Manajemen Indonesia*, 03(03).

Parra, F., Jacobs, A., & Trevino, L. L. (2021). Shippy Express: Augmenting accounting education with Google Sheets. *Journal of Accounting Education*, 56, 100740.

Pemerintah Indonesia. (2009). Undang-Undang (UU) Nomor 10 Tahun 2009 tentang Kepariwisataan. Indonesia: Pemerintah Republik Indonesia.

Prasetyo, Achmad dan Sukim. (2016). Modul Metode Pengumpulan Data Program Studi Statistika Program Diploma III. Jakarta: Sekolah Tinggi Ilmu Statistik (STIS).

Pratama, R. R. (n.d.). *Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia* (Vol. 19 No. 2, 2020, hlm. 302-311). Jurnal MATRIK.

Putra, D., Daniawan, B., Suwitno, S., & Wijaya, A. (2019). The Analysis and Design Marketplace Information Systems Web-Based of Electronic Repair Service Providers with Haversine Method. *bit-Tech*, 2(1), 53-62.

Putra, A. P. (2021). *AMDA: Anchor Mobility Data Analytic for Determining Home-Work Location from Mobile Positioning Data* (Vol. 2021 No. 1, 2021, hlm.296). ICDSOS.

Putra, A. P., Munaf, A. R., Ruslani, A. (2021). Combination methods in tourism statistics. In CONFERENCE OF EUROPEAN STATISTICIANS: *Expert Meeting on Statistical Data Collection*: UNECE.

Putra, A. P., Madjida, W. O. Z., Setyadi, I. A., Nugroho, A. R. S., & Munaf, A. R. M. (2021). AMDA: Anchor Mobility Data Analytic for Determining Home-Work Location from Mobile Positioning Data. In *Proceedings of The International Conference on Data Science and Official Statistics* (Vol. 2021, No. 1, pp. 296-304).

Raihan rafiful Allaam, M., & Wibowo, A. T. (2021). Klasifikasi Genus Tanaman Anggrek Menggunakan Metode Convolutional Neural Network (CNN). *E-Prceeding of Engineering*, 8(2), 1–1153.

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Science Direct*, 57(Tourism Management), 202-212. <https://doi.org/10.1016/j.tourman.2016.06.006>

Retnoningsih, E., & Pramudita, R. (n.d.). *Mengenal Machine Learning dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python* (Vol. 7 No. 2, 2020, hlm.156-165). Bina Insani ICT Journal.

Rohan, A. (n.d.). *Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper* (Vol. 5 No.1, 2019, hlm.75-82). IJCIT.

Sedqui, A. (1995). *Manufacturing Squence Family Grouping for FMS Design - A New Approach* (Vol. 1995 hlm. 429-437).

Semczuk, N. (2020). 4 Times You Should Really Be Using Google Sheets Instead of Excel. <https://www.themuse.com/advice/4-times-you-should-really-be-using-google-sheets-instead-of-excel>

Senin, P. (2008). Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23), 40.

Singh, J. (n.d.). *Problem of Double Counting: Meaning and Ways for Avoiding Micro Economics*. <https://www.economicsdiscussion.net/national-income/problem-of-double-counting-meaning-and-ways-for-avoiding-micro-economics/499>

Sugiyono. (2005). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. In Metode Penelitian. Bandung: Alfabeta

Suthaharan, S. (2016). *Decision Tree Learning. Integrated Series in Information Systems* (Vol. 237–269). 10.1007/978-1-4899-7641-3_10

Taylor, K., & Sakyi. (2016). *Big Data: Understanding Big Data*.

Tillman, K. H., & Nam, C. B. (2008). *Family Structure Outcomes of Alternative Family Definitions* (Vol. 27, 2008, hlm. 367-384). Springer Science+Business Media B.V.

Toohey, K., & Duckham, M. (2015). Trajectory similarity measures. *Sigspatial Special*, 7(1), 43-50.

Uber Technologies, I. (2023). *H3 Hexagonal hierarchical geospatial indexing system*. <https://h3geo.org/>



UNWTO. (2008). *Glossary Of Tourism Terms*.

Vallat, R. (2018). *Pingouin: statistics in Python* (Vol. 2018 hlm. 1). Berkeley : University of California.

Walpole, R. E. (1995). Pengantar Statistika. Jakarta: Penerbit PT Gramedia Pustaka Utama.

The World Tourism Organization. (2010). *International Recommendations for Tourism Statistics 2008*. UNWTO.
<https://www.e-unwto.org/doi/epdf/10.18111/9789211615210>

Yaqoob et al., I. (2016). *Big Data: From beginning to future, (International Jurnal of Information Management, 2016)*, 1231-1247.

Zaragozi, B., Trilles, S., & Gutierrez, A. (2021). Passive mobile data for studying seasonal tourism mobilities: An application in a Mediterranean coastal destination. *ISPRS International Journal of Geo-Information*, 10(2), 98.

Zhi, T., Luo, H., & Liu, Y. (2018). *A Gini impurity-based interest flooding attack defence mechanism in NDN* (Vol. 22(3), 538-541). IEEE Communications Letters.

Zhou, J., Ben, J., Wang, R., Zheng, M., & Du, L. (2020). Lattice quad-tree indexing algorithm for a hexagonal discrete global grid system. *ISPRS International Journal of Geo-Information*, 9(2), 83.

Zicari, R. (2013). *Big Data: Challenges and Opportunities*.



LAMPIRAN

Lampiran 1. Tabulasi Hasil *Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint* untuk *Family Grouping*

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,25	0,40	36
False FG	0,92	1,00	0,96	319
Macro Avg	0,96	0,62	0,68	355
Accuracy			0,92	

Lampiran 2. Tabulasi Hasil *Running Algoritma H3 Tanpa Pembentukan Staypoint* untuk *Family Grouping*

Resolusi 1

Kelas	Precision	Recall	F1-Score	Support
True FG	0,86	0,17	0,28	36
False FG	0,91	1,00	0,95	319
Macro Avg	0,89	0,58	0,62	355
Accuracy			0,91	

Resolusi 2

Kelas	Precision	Recall	F1-Score	Support
True FG	0,50	0,23	0,31	35
False FG	0,92	0,97	0,95	319
Macro Avg	0,71	0,60	0,63	354
Accuracy			0,90	



Resolusi 3

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,03	0,06	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,95	0,51	0,50	354
Accuracy			0,90	

Resolusi 4

Kelas	Precision	Recall	F1-Score	Support
True FG	0,00	0,00	0,00	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,45	0,50	0,47	354
Accuracy			0,90	

Resolusi 5

Kelas	Precision	Recall	F1-Score	Support
True FG	0,00	0,00	0,00	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,45	0,50	0,47	354
Accuracy			0,90	

Resolusi 6

Kelas	Precision	Recall	F1-Score	Support
True FG	0,50	0,03	0,05	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,70	0,51	0,50	354
Accuracy			0,90	

Resolusi 7

Kelas	Precision	Recall	F1-Score	Support
True FG	0,50	0,03	0,05	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,70	0,51	0,50	354
Accuracy			0,90	

Resolusi 8

Kelas	Precision	Recall	F1-Score	Support
True FG	0,00	0,00	0,00	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,45	0,50	0,47	354
Accuracy			0,90	

Resolusi 9

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,03	0,06	35
False FG	0,90	1,00	0,95	319
Macro Avg	0,95	0,51	0,50	354
Accuracy			0,90	

Resolusi 10

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,08	0,15	36
False FG	0,91	1,00	0,95	319
Macro Avg	0,95	0,54	0,55	355
Accuracy			0,91	



Resolusi 11

Kelas	Precision	Recall	F1-Score	Support
True FG	0,00	0,00	0,00	36
False FG	0,90	1,00	0,95	319
Macro Avg	0,45	0,50	0,47	355
Accuracy				0,90

Resolusi 12

Kelas	Precision	Recall	F1-Score	Support
True FG	0,00	0,00	0,00	36
False FG	0,90	1,00	0,95	319
Macro Avg	0,45	0,50	0,47	355
Accuracy				0,90

Lampiran 3. Tabulasi Hasil *Running* Algoritma H3 dengan Pembentukan *Staypoint* untuk *Family Grouping*

Resolusi 1

Kelas	Precision	Recall	F1-Score	Support
True FG	0,07	0,53	0,13	36
False FG	0,82	0,25	0,38	319
Macro Avg	0,45	0,39	0,26	355
Accuracy				0,28



Resolusi 2

Kelas	Precision	Recall	F1-Score	Support
True FG	0,24	0,22	0,23	36
False FG	0,91	0,92	0,92	319
Macro Avg	0,58	0,57	0,57	355
Accuracy				0,85

Resolusi 3

Kelas	Precision	Recall	F1-Score	Support
True FG	0,55	0,44	0,49	36
False FG	0,94	0,96	0,95	319
Macro Avg	0,75	0,70	0,72	355
Accuracy				0,91

Resolusi 4

Kelas	Precision	Recall	F1-Score	Support
True FG	0,61	0,47	0,53	36
False FG	0,94	0,97	0,95	319
Macro Avg	0,77	0,72	0,74	355
Accuracy				0,92

Resolusi 5

Kelas	Precision	Recall	F1-Score	Support
True FG	0,51	0,61	0,56	36
False FG	0,96	0,93	0,94	319
Macro Avg	0,73	0,77	0,75	355
Accuracy				0,90



Resolusi 6

Kelas	Precision	Recall	F1-Score	Support
True FG	0,88	0,64	0,74	36
False FG	0,96	0,99	0,98	319
Macro Avg	0,92	0,81	0,86	355
Accuracy			0,95	

Resolusi 7

Kelas	Precision	Recall	F1-Score	Support
True FG	0,92	0,61	0,73	36
False FG	0,96	0,99	0,98	319
Macro Avg	0,94	0,80	0,85	355
Accuracy			0,95	

Resolusi 8

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,67	0,80	36
False FG	0,96	1,00	0,98	319
Macro Avg	0,98	0,83	0,89	355
Accuracy			0,97	

Resolusi 9

Kelas	Precision	Recall	F1-Score	Support
True FG	0,97	0,78	0,86	36
False FG	0,98	1,00	0,99	319
Macro Avg	0,97	0,89	0,92	355
Accuracy			0,97	

Resolusi 10

Kelas	Precision	Recall	F1-Score	Support
True FG	0,97	0,81	0,88	36
False FG	0,98	1,00	0,99	319
Macro Avg	0,97	0,90	0,93	355
Accuracy				0,98

Resolusi 11

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,42	0,59	36
False FG	0,94	1,00	0,97	319
Macro Avg	0,97	0,71	0,78	355
Accuracy				0,94

Resolusi 12

Kelas	Precision	Recall	F1-Score	Support
True FG	1,00	0,42	0,59	36
False FG	0,94	1,00	0,97	319
Macro Avg	0,97	0,71	0,78	355
Accuracy				0,94

Lampiran 4. Tabulasi Hasil *Running Dynamic Time Warping* untuk *Family Grouping*

Kelas	Precision	Recall	F1-Score	Support
True FG	0,25	0,03	0,05	36
False FG	0,90	0,99	0,94	319
Macro Avg	0,58	0,51	0,50	355
Accuracy				0,89



Lampiran 5. Tabulasi Hasil *Running Frechet Distance* untuk *Family Grouping*

Kelas	Precision	Recall	F1-Score	Support
True FG	0,15	0,06	0,08	36
False FG	0,90	0,97	0,93	319
Macro Avg	0,53	0,51	0,51	355
Accuracy				0,87

Lampiran 6. Tabulasi Hasil *Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint* untuk *Double Counting* Periode Bulanan

Kelas	Precision	Recall	F1-Score	Support
True DC	0,58	0,54	0,56	13
False DC	0,99	0,99	0,99	406
Macro Avg	0,78	0,76	0,77	419
Accuracy				0,97

Lampiran 7. Tabulasi Hasil *Running Algoritma Kombinasi Jarak Terdekat Antar Staypoint* untuk *Double Counting* Periode Tahunan

Kelas	Precision	Recall	F1-Score	Support
True DC	0,60	0,69	0,64	13
False DC	0,99	0,99	0,99	406
Macro Avg	0,80	0,84	0,82	419
Accuracy				0,98



Lampiran 8. Tabulasi Hasil *Running Algoritma H3 Tanpa Pembentukan Staypoint* untuk *Double Counting*

Resolusi 1

Kelas	Precision	Recall	F1-Score	Support
True DC	0,54	0,54	0,54	13
False DC	0,99	0,99	0,99	406
Macro Avg	0,76	0,76	0,76	419
Accuracy				0,97

Resolusi 2

Kelas	Precision	Recall	F1-Score	Support
True DC	0,21	0,31	0,25	13
False DC	0,98	0,96	0,97	406
Macro Avg	0,59	0,64	0,61	419
Accuracy				0,94

Resolusi 3

Kelas	Precision	Recall	F1-Score	Support
True DC	0,36	0,31	0,33	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,67	0,65	0,66	419
Accuracy				0,96

Resolusi 4

Kelas	Precision	Recall	F1-Score	Support
True DC	0,36	0,31	0,33	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,67	0,65	0,66	419
Accuracy				0,96



Resolusi 5

Kelas	Precision	Recall	F1-Score	Support
True DC	0,67	0,31	0,42	13
False DC	0,98	1,00	0,99	406
Macro Avg	0,82	0,65	0,70	419
Accuracy	0,97			

Resolusi 6

Kelas	Precision	Recall	F1-Score	Support
True DC	0,80	0,31	0,44	13
False DC	0,98	1,00	0,99	406
Macro Avg	0,89	0,65	0,72	419
Accuracy	0,98			

Resolusi 7

Kelas	Precision	Recall	F1-Score	Support
True DC	0,57	0,31	0,40	13
False DC	0,98	0,99	0,99	406
Macro Avg	0,77	0,65	0,69	419
Accuracy	0,97			

Resolusi 8

Kelas	Precision	Recall	F1-Score	Support
True DC	0,57	0,31	0,40	13
False DC	0,98	0,99	0,99	406
Macro Avg	0,77	0,65	0,69	419
Accuracy	0,97			



Resolusi 9

Kelas	Precision	Recall	F1-Score	Support
True DC	0,40	0,31	0,35	13
False DC	0,98	0,99	0,98	406
Macro Avg	0,69	0,65	0,66	419
Accuracy	0,96			

Resolusi 10

Kelas	Precision	Recall	F1-Score	Support
True DC	0,27	0,23	0,25	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,62	0,61	0,61	419
Accuracy	0,96			

Resolusi 11

Kelas	Precision	Recall	F1-Score	Support
True DC	0,30	0,23	0,26	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,64	0,61	0,62	419
Accuracy	0,96			

Resolusi 12

Kelas	Precision	Recall	F1-Score	Support
True DC	0,27	0,31	0,29	13
False DC	0,98	0,97	0,98	406
Macro Avg	0,62	0,64	0,63	419
Accuracy	0,95			



Lampiran 9. Tabulasi Hasil *Running* Algoritma H3 dengan Pembentukan *Staypoint* untuk *Double Counting*

Resolusi 1

Kelas	Precision	Recall	F1-Score	Support
True DC	0,20	0,08	0,11	13
False DC	0,97	0,99	0,98	406
Macro Avg	0,59	0,53	0,55	419
Accuracy			0,96	

Resolusi 2

Kelas	Precision	Recall	F1-Score	Support
True DC	0,00	0,00	0,00	13
False DC	0,97	0,99	0,98	406
Macro Avg	0,48	0,50	0,49	419
Accuracy			0,96	

Resolusi 3

Kelas	Precision	Recall	F1-Score	Support
True DC	0,00	0,00	0,00	13
False DC	0,97	0,98	0,97	406
Macro Avg	0,48	0,49	0,49	419
Accuracy			0,95	

Resolusi 4

Kelas	Precision	Recall	F1-Score	Support
True DC	0,27	0,23	0,25	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,62	0,61	0,61	419
Accuracy			0,96	



Resolusi 5

Kelas	Precision	Recall	F1-Score	Support
True DC	0,26	0,38	0,31	13
False DC	0,98	0,97	0,97	406
Macro Avg	0,62	0,68	0,64	419
Accuracy	0,95			

Resolusi 6

Kelas	Precision	Recall	F1-Score	Support
True DC	0,26	0,38	0,31	13
False DC	0,98	0,97	0,97	406
Macro Avg	0,62	0,68	0,64	419
Accuracy	0,95			

Resolusi 7

Kelas	Precision	Recall	F1-Score	Support
True DC	0,25	0,23	0,24	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,61	0,60	0,61	419
Accuracy	0,95			

Resolusi 8

Kelas	Precision	Recall	F1-Score	Support
True DC	0,43	0,46	0,44	13
False DC	0,98	0,98	0,98	406
Macro Avg	0,71	0,72	0,71	419
Accuracy	0,96			



Resolusi 9

Kelas	Precision	Recall	F1-Score	Support
True DC	0,40	0,31	0,35	13
False DC	0,98	0,99	0,98	406
Macro Avg	0,69	0,65	0,66	419
Accuracy	0,96			

Resolusi 10

Kelas	Precision	Recall	F1-Score	Support
True DC	0,38	0,23	0,29	13
False DC	0,98	0,99	0,98	406
Macro Avg	0,68	0,61	0,63	419
Accuracy	0,96			

Resolusi 11

Kelas	Precision	Recall	F1-Score	Support
True DC	0,12	0,08	0,10	13
False DC	0,97	0,98	0,98	406
Macro Avg	0,55	0,53	0,54	419
Accuracy	0,95			

Resolusi 12

Kelas	Precision	Recall	F1-Score	Support
True DC	0,11	0,08	0,09	13
False DC	0,97	0,98	0,98	406
Macro Avg	0,54	0,53	0,53	419
Accuracy	0,95			



Lampiran 10. Tabulasi Hasil *Running Dynamic Time Warping* untuk *Double Counting*

Kelas	Precision	Recall	F1-Score	Support
True DC	0,00	0,00	0,00	13
False DC	0,97	1,00	0,98	406
Macro Avg	0,48	0,50	0,49	419
Accuracy				0,96

Lampiran 11. Tabulasi Hasil *Running Frechet Distance* untuk *Double Counting*

Kelas	Precision	Recall	F1-Score	Support
True DC	0,00	0,00	0,00	13
False DC	0,97	1,00	0,98	406
Macro Avg	0,48	0,50	0,49	419
Accuracy				0,97



Lampiran 12. Kuesioner Survei Kajian MPD (SPRT.PKL62)

 STIS	PRAKTIK KERJA LAPANGAN POLITEKNIK STATISTIKA STIS SURVEI RUMAH TANGGA PENGGUNA TELKOMSEL DAN INDOSAT	 SPRT.PKL62
RAHASIA		
100. a Jam Mulai Wawancara : WIB 100. b Jam Selesai Wawancara : WIB		
BLOK I. IDENTITAS WILAYAH		
101	Provinsi	JAWA TIMUR <input type="checkbox"/> 3 <input type="checkbox"/> 5
102	Kabupaten/Kota*	<input type="checkbox"/> <input type="checkbox"/>
103	Kecamatan	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
104	Desa/Kelurahan*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
105	Nama Kepala Rumah Tangga
106	Alamat (nama jalan/gang, RT/RW/Dusun)
107	Nama pemberi informasi
BLOK II. RINGKASAN		
201. Banyak anggota rumah tangga		
202. Banyak anggota rumah tangga yang memiliki kartu sim aktif		
203. Banyak anggota rumah tangga yang menggunakan 2 kartu sim Telkomsel dalam 1 handphone		
204. Banyak anggota rumah tangga yang menggunakan 2 kartu sim Indosat dalam 1 handphone		
205. Banyak anggota rumah tangga yang menggunakan kartu sim Telkomsel dan Indosat secara bersama dalam 1 handphone		
*) Coret yang tidak perlu		
BLOK IIIA. KETERANGAN PETUGAS		
Uraian (1)	Nama (2)	NIM (3)
301. PPL	Tanggal (4) tgl <input type="checkbox"/> bin <input type="checkbox"/>
302. PML	Kode Tim (5) tgl <input type="checkbox"/> bln <input type="checkbox"/>
Tanda Tangan (6)		
BLOK IIIB. KETERANGAN PENCACAHAN		
303. Hasil Pencacahan 1. Berhasil diwawancara 2. Tidak dapat diwawancara sampai dengan batas waktu pencacahan 3. Menolak diwawancara (.....)		
BLOK IV. CATATAN		





BLOK VIII. KETERANGAN PERJALANAN TERAKHIR YANG DILAKUKAN RUMAH TANGGA	
Pertanyaan 801 sampai 805 merupakan rincian perjalanan terakhir yang dilakukan oleh minimal 2 ART relawan yang dilakukan dalam periode 1 Januari 2022 - 31 Desember 2022	
801. Nam anggota rumah tangga : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/>	805. Apa daerah tujuan utama yang dikunjungi : <input type="checkbox"/> <input type="checkbox"/>
802. Kapan perjalanan dilakukan : Bulan <input type="checkbox"/> <input type="checkbox"/> Minggu ke- <input type="checkbox"/> <input type="checkbox"/> [Isikan salah satu kode berikut dalam kotak "Minggu ke-"] (1) Tanggal 1 - 7 (2) Tanggal 8 - 14 (3) Tanggal 15 - 21 (4) Tanggal 22 - 31	a. Provinsi : <input type="checkbox"/> <input type="checkbox"/> b. Kab.Kota : <input type="checkbox"/> <input type="checkbox"/> c. Kecamatan : <input type="checkbox"/> <input type="checkbox"/>
803. Apa moda angkutan utama yang digunakan dalam rangka melakukan perjalanan: (1) Pesawat (2) Kapal laut / angkutan air <input type="checkbox"/> (3) Kereta api (4) Kendaraan roda 4 / bus (5) Sepeda motor (6) Lainnya, sebutkan (.....)	807. Siapa saja ART relawan yang mengikuti perjalanan: a. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> b. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> c. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> d. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> e. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> f. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> g. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/> h. Nama : No. Urut ART : <input type="checkbox"/> <input type="checkbox"/>
804. Berapa lama durasi perjalanan <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> jam	
805. Apa daerah asal perjalanan :	
a. Provinsi : <input type="checkbox"/> <input type="checkbox"/> b. Kab.Kota : <input type="checkbox"/> <input type="checkbox"/> c. Kecamatan : <input type="checkbox"/> <input type="checkbox"/>	

Lampiran 13. Kuesioner Survei Wisatawan Nusantara (VRTW.62.C)

PRAKTIK KERJA LAPANGAN D-IV T.A. 2022/2023 POLITEKNIK STATISTIKA STIS PILOT SURVEI WISATAWAN NUSANTARA				VRTW.62.C
	RAHASIA	Jam Mulai Wawancara		WIB
		Jam Selesai Wawancara		WIB
Kode Sampel Ruta Terpilih: 1 = Riset 1 (pertanyaan * dan **) 2 = Riset 2 (pertanyaan ** dan ***) 3 = Riset 1 dan 2 (seluruh pertanyaan)				
Keterangan: * = Ditanyakan untuk riset 1 saja ** = Ditanyakan untuk riset 2 saja *** = Ditanyakan untuk riset 1 dan 2				
BLOK I. KETERANGAN TEMPAT RUMAH TANGGA				
(1)	(2)	(3)	(4)	
101	Provinsi***	Jawa Timur	<input type="checkbox"/> 3	<input type="checkbox"/> 5
102	Kota***	Malang	<input type="checkbox"/> 7	<input type="checkbox"/> 3
103	Kecamatan***	<input type="checkbox"/>	<input type="checkbox"/>
104	Desa/Kelurahan***	<input type="checkbox"/>	<input type="checkbox"/>
105	Nomor Blok Sensus***		<input type="checkbox"/>	<input type="checkbox"/>
106	Nomor Kode Sampel***		<input type="checkbox"/>	<input type="checkbox"/>
107	Nomor Urut Rumah Tangga Jalan***		<input type="checkbox"/>	<input type="checkbox"/>
108	Keterangan Listing***	a. Pemberi Informasi Utama [nama] b. Pelaku Perjalanan [nama] c. Tujuan Perjalanan [sian]		
109	Nama Kepala Rumah Tangga***		
110	Alamat (Nama Jalan/Gang, RT/RW/Dusun)***			
111	Kendaraan Bermotor yang Dimiliki Rumah Tangga**	Mobil: <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Sepeda Motor: <input type="checkbox"/> <input type="checkbox"/>		
Kode 103: Kode 104: Kode 105: Kode 106: Kode 107: Kode 108: Kode 109: Kode 110: 010 Kecamatan Kedungkandang 009 Swejajar 005 Bandungrejojati 002 Subuharjo 010 Arjosari 007 Tukorejo 010 Kecamatan 010 Madyapuro 006 Bukanul Krajan 003 Kulihalem 011 Baléarjosari 008 Mojolenggu Kedungkandang 011 Losarpuro 007 Muylerejo 004 Kauman 012 Jodjan 009 Turungekar 020 Kecamatan Sukun 012 Cemorokandang 008 Bandulan 005 Bareng 003 Keluhan 030 Kecamatan Klojen 001 Arjowinangun 009 Tanjungrejo 006 Gadingkastri 003 Kesatrian 040 Kecamatan Blimbing 002 Tlogowaru 010 Pisang Candi 007 Oro Oro Dowo 010 Taskmadu 050 Kecamatan Lowokwaru 003 Wonokoyo 001 Kebonsari 008 Klojen 004 Buncurejo 011 Tungulwulung 060 Kecamatan Kedungkandang 004 Banjireuy 011 Karang Besuki 005 Purwantoro 002 Dinoyo 060 Kecamatan 005 Banting 003 Optomulyo 009 Rampal Celaket 003 Sumberasri 065 Merbabu 006 Merbabu 001 Kasin 010 Samaan 007 Blimbing 004 Ketawanggede 007 Ketelan 008 Kedungkandang 011 Penanggungan 011 Peranggungan 008 Purvodadi 009 Kedungkandang 009 Kedungkandang 012 Polowijen 005 Jatimulyo 006 Lowokwaru				

BLOK II. RINGKASAN			
(1)	(2)	(3)	
201	Banyaknya anggota rumah tangga [Nomor urut terakhir pada rincian kolom 501]***	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
202	Banyaknya anggota rumah tangga yang melakukan "Perjalanan" selama periode: Oktober 2022-Desember 2022 [Jumlah ART yang mengisi pertanyaan 523 s.d. 528 bernilai 1 (Ya)]*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
203	Banyaknya anggota rumah tangga yang melakukan "Perjalanan di dalam kota" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang menjawab minimal satu kolom dari 523 s.d. 525 bernilai 1 (Ya)]**	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
203a	Banyaknya anggota rumah tangga yang melakukan "Perjalanan di dalam kota menuju ke objek wisata" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 523 = 1]**	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
203b	Banyaknya anggota rumah tangga yang melakukan "Perjalanan di dalam kota ke akomodasi komersial" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 524 = 1]**	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
203c	Banyaknya anggota rumah tangga yang melakukan "Perjalanan di dalam kota untuk tujuan lainnya" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 525 = 1]**	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
204	Banyaknya anggota rumah tangga yang melakukan "Perjalanan ke luar kota" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang menjawab minimal satu kolom dari 526 s.d. 528 bernilai 1 (Ya)]*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
204a	Banyaknya anggota rumah tangga yang melakukan "Perjalanan ke luar kota menuju ke objek wisata" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 526 = 1]*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
204b	Banyaknya anggota rumah tangga yang melakukan "Perjalanan ke luar kota menuju ke akomodasi komersial" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 527 = 1]*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
204c	Banyaknya anggota rumah tangga yang melakukan "Perjalanan ke luar kota untuk tujuan lainnya" selama periode: Oktober 2022-Desember 2022 [Banyak ART yang mengisi pertanyaan 528 = 1]*	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	



BLOK III. KETERANGAN PETUGAS					
Uraian (1)	Nama (2)	NIM (3)	Tanggal (4)	Nomor Tim (5)	Tanda Tangan (6)
301 PPL***	Tgl <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
302 PML***	Tgl <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

BLOK IV. CATATAN



Riset I

Laporan Hasil

Praktik Kerja Lapangan D-IV
Politeknik Statistika STIS
T.A. 2022/2023

BLOK VI. RINCIAN BANYAK PERJALANAN											
(601) No. Urut ART*** <input type="checkbox"/> <input type="checkbox"/>	Ditanyakan jika jawaban ART pada pertanyaan [kol 523 s.d. 525 minimal satu kolom bernilai ("1"=Ya)]										
	Berapa kali [nama] melakukan perjalanan di dalam kota tidak lebih dari satu tahun dan telah kembali pada masing-masing bulan berikut (Oktober 2022-Desember 2022)***?			Kecamatan Kedungkandang			Kecamatan Sukun		Kecamatan Klojen		
Bulan	Ke Objek Wisata	Ke Akomodasi Komersial	Selain Itu	Ke Objek Wisata	Ke Akomodasi Komersial	Selain Itu	Ke Objek Wisata	Ke Akomodasi Komersial	Selain Itu		
	(602)	(603)	(604)	(605)	(606)	(607)	(608)	(609)	(610)	(611)	
	Oktober	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	
	November	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	
	Desember	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	
	Total Perjalanan	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	
	Total Perjalanan per kecamatan	<input type="checkbox"/> <input type="checkbox"/>			<input type="checkbox"/> <input type="checkbox"/>			<input type="checkbox"/> <input type="checkbox"/>			
	(601) No. Urut ART*** <input type="checkbox"/> <input type="checkbox"/>	Ditanyakan jika jawaban ART pada pertanyaan [kol 523 s.d. 525 minimal satu kolom bernilai ("1"=Ya)]									Ditanyakan jika jawaban ART pada pertanyaan [kol 526 s.d. 528 minimal satu kolom bernilai ("1"=Ya)]
		Berapa kali [nama] melakukan perjalanan di dalam kota tidak lebih dari satu tahun dan telah kembali pada masing-masing bulan berikut (Oktober 2022-Desember 2022)***?									Berapa kali [nama] melakukan perjalanan ke luar kota tidak lebih dari satu tahun dan telah kembali pada masing-masing bulan berikut (Oktober 2022-Desember 2022)*?
		Bulan	Kecamatan Blimbing			Kecamatan Lowokwaru					
Ke Objek Wisata			Ke Akomodasi Komersial	Selain Itu	Ke Objek Wisata	Ke Akomodasi Komersial	Selain Itu	Ke Objek Wisata	Ke Akomodasi Komersial	Selain Itu	
(602)			(612)	(613)	(614)	(615)	(616)	(617)	(618)	(619)	(620)
Oktober			<input type="checkbox"/> <input type="checkbox"/>								
November			<input type="checkbox"/> <input type="checkbox"/>								
Desember			<input type="checkbox"/> <input type="checkbox"/>								
Total Perjalanan			<input type="checkbox"/> <input type="checkbox"/>								
Total Perjalanan per kecamatan			<input type="checkbox"/> <input type="checkbox"/>			<input type="checkbox"/> <input type="checkbox"/>			<input type="checkbox"/> <input type="checkbox"/>		

BLOK IX RINCIAN BIAYA PERJALANAN ROMBONGAN			
Ditanyakan Jika 710 = ("2")			
Berapa Rupiah Biaya Aktivitas Berikut selama Perjalanan Tersebut?			
Termasuk yang dibayar oleh pihak lain (reimburse). Jika tidak ada, Isikan 0			
907. Paket Perjalanan**		911. Belanja**	
a. Biaya paket perjalanan rombongan (Rp)		a. Biaya belanja rombongan (Rp)	
b. Biaya tambahan yang digunakan oleh [nama] (Rp)		b. Biaya tambahan yang digunakan oleh [nama] (Rp)	
c. Biaya per kapita (Rp)		c. Biaya per kapita (Rp)	
908. Pramuwisata**		912. Kesehatan**	
a. Biaya pramuwisata rombongan (Rp)		a. Biaya kesehatan rombongan (Rp)	
b. Biaya tambahan yang digunakan oleh [nama] (Rp)		b. Biaya tambahan yang digunakan oleh [nama] (Rp)	
c. Biaya per kapita (Rp)		c. Biaya per kapita (Rp)	
909. Hiburan**		913. Lainnya**	
a. Biaya hiburan rombongan (Rp)		a. Biaya lainnya rombongan (Rp)	
b. Biaya tambahan yang digunakan oleh [nama] (Rp)		b. Biaya tambahan yang digunakan oleh [nama] (Rp)	
c. Biaya per kapita (Rp)		c. Biaya per kapita (Rp)	
910. Cinderamata**		914. Total Pengeluaran**	
a. Biaya cinderamata rombongan (Rp)		Total biaya per kapita (Rp)	
b. Biaya tambahan yang digunakan oleh [nama] (Rp)			
c. Biaya per kapita (Rp)			

Definisi Biaya

- Biaya tambahan yang digunakan oleh [nama] adalah biaya yang digunakan untuk kepentingan pribadi oleh individu tersebut dalam perjalanan rombongan.
- Biaya per kapita = (biaya rombongan/jumlah anggota rombongan) + biaya tambahan yang digunakan oleh [nama].



Lampiran 14. Agreement Relawan Kajian MPD Provider Telkomsel

**Formulir Persetujuan Penggunaan Data Pribadi Relawan
untuk Mobile Positioning Data Wisatawan Nusantara 2022**

Saya yang bertandatangan di bawah ini:

Nama : ...

Nomor Identitas KTP : ...

Nomor HP (Telkomsel) : ...

Dengan penuh kesadaran dan tanpa paksaan, bersedia secara sukarela untuk memberikan informasi dan data pribadi saya pada periode Desember 2020 sampai Desember 2022, meliputi:

1. Nomor *handphone* (HP) Telkomsel yang masih aktif pada periode Desember 2020 s.d. Desember 2022.
2. Tanggal dan waktu transaksi yang ter-*capture* di sistem Telkomsel beserta ID provinsi, kabupaten/kota, kecamatan dan desa/kelurahan tempat terjadinya transaksi tersebut, termasuk perkiraan koordinat lokasi terjadinya transaksi.

untuk selanjutnya disebut sebagai "Data Pribadi".

Dengan menandatangani formulir persetujuan ini, saya memberikan wewenang dan kuasa penuh kepada Telkomsel dan Badan Pusat Statistik (BPS) untuk:

1. Memperoleh, mengumpulkan, mengolah, menganalisa, menyimpan, menampilkan, mengumumkan, mengirimkan, menyebarkan dan/atau memusnahkan Data Pribadi hanya untuk keperluan penjaminan kualitas data MPD Wisnus.
2. Menyimpan Data Pribadi pada server BPS dan memastikan keamanan data melalui proses enkripsi dan/atau bentuk lainnya yang bertujuan untuk perlindungan data.

Dengan menandatangani formulir ini dan memberikan persetujuan di atas, saya melepasskan tanggung jawab Telkomsel sebagai pengelola dan pengolah Data Pribadi, dalam hal terdapat aktivitas lain yang dilakukan terhadap Data Pribadi selain dari poin persetujuan di atas.

Persetujuan ini berlaku sejak tanggal persetujuan ini ditandatangani sepanjang tidak terdapat keberatan secara tertulis untuk menghentikan persetujuan ini.

Yang memberikan persetujuan,

Nama: ...

Materai 10.000,-
ttd

Tanggal:



Lampiran 15. *Agreement Relawan Kajian MPD Provider Indosat*

**Formulir Persetujuan Penggunaan Data Pribadi Relawan
untuk Mobile Positioning Data Wisatawan Nusantara 2022**

Saya yang bertandatangan di bawah ini:

Nama : ...

Nomor Identitas KTP : ...

Nomor HP (Indosat) : ...

Dengan penuh kesadaran dan tanpa paksaan, bersedia secara sukarela untuk memberikan informasi dan data pribadi saya pada periode Desember 2020 sampai Desember 2022, meliputi:

1. Nomor handphone (HP) Indosat yang masih aktif pada periode Desember 2022 s.d. Desember 2022.
2. Tanggal dan waktu transaksi yang ter-capture di sistem Indosat beserta ID provinsi, kabupaten/kota, kecamatan dan desa/kelurahan tempat terjadinya transaksi tersebut, termasuk perkiraan koordinat lokasi terjadinya transaksi.

untuk selanjutnya disebut sebagai "Data Pribadi".

Dengan menandatangani formulir persetujuan ini, saya memberikan wewenang dan kuasa penuh kepada Indosat dan Badan Pusat Statistik (BPS) untuk:

1. Memperoleh, mengumpulkan, mengolah, menganalisa, menyimpan, menampilkan, mengumumkan, mengirimkan, menyebarluaskan dan/atau memusnahkan Data Pribadi hanya untuk keperluan penjaminan kualitas data MPD Wisnus.
2. Menyimpan Data Pribadi pada server BPS dan memastikan keamanan data melalui proses enkripsi dan/atau bentuk lainnya yang bertujuan untuk perlindungan data.

Dengan menandatangani formulir ini dan memberikan persetujuan di atas, saya melepaskan tanggung jawab Indosat sebagai pengelola dan pengolah Data Pribadi, dalam hal terdapat aktivitas lain yang dilakukan terhadap Data Pribadi selain dari poin persetujuan di atas.

Persetujuan ini berlaku sejak tanggal persetujuan ini ditandatangani sepanjang tidak terdapat keberatan secara tertulis untuk menghentikan persetujuan ini.

Yang memberikan persetujuan,

Nama: ...

Materai 10.000,-

Ttd

Tanggal:



**Bersinergi dengan Data,
Majukan Pariwisata Nusantara**

