SKRIPSI

SEGMENTASI PENGUNJUNG WISATA BERDASARKAN *WILLINGNESS TO PAY* BERBASIS WEBSITE MENGGUNAKAN MODEL *KMEANS* (STUDI KASUS TAMAN WISATA ALAM LEJJA)

Disusun dan diajukan oleh:

WIRA DRANA WASISTHA
D121 19 1017

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024

LEMBAR PENGESAHAN SKRIPSI

SEGMENTASI PENGUNJUNG WISATA BERDASARKAN *WILLINGNESS TO PAY* BERBASIS WEBSITE MENGGUNAKAN MODEL *KMEANS* (STUDI KASUS TAMAN WISATA ALAM LEJJA)

Disusun dan diajukan oleh

Wira Drana Wasistha
D121 19 1017

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 07 Mei 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM. ASEAN. Eng.
NIP 197507162002121004

Pembimbing Pendamping,

Elly Warni, S.T., M.T.
NIP 198202162008122001

Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM. ASEAN. Eng.
NIP 197507162002121004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini;

Nama	:	Wira Drana Wasistha
NIM	:	D121191017
Program Studi	:	Teknik Informatika
Jenjang	:	S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

SEGMENTASI PENGUNJUNG WISATA BERDASARKAN *WILLINGNESS TO PAY* BERBASIS WEBSITE MENGGUNAKAN MODEL *KMEANS* (STUDI KASUS TAMAN WISATA ALAM LEJJA)

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 07 Mei 2024

ng Menyatakan

Wira Drana Wasistha

ABSTRAK

WIRA DRANA WASISTHA. *Segmentasi Pengunjung Wisata Berdasarkan Willingness to Pay Berbasis Website Menggunakan Model KMeans (Studi Kasus Taman Wisata Alam Lejja)* (dibimbing oleh Indrabayu dan Elly Warni)

Pengelola Taman Wisata Alam (TWA) Lejja telah melakukan penelitian tentang kesediaan membayar (*Willingness to Pay*/*WTP*) pengunjung di kawasan konservasi untuk perbaikan dan penambahan fasilitas umum dasar, seperti gerbang masuk, jaringan jalan, *Community Empowerment Area* (*CEA*), toilet, masjid/mushola, tempat parkir, serta gazebo/*shelter* . Pengelola memerlukan penelitian lanjutan untuk menganalisis segmentasi pengunjung berdasarkan *WTP* untuk menentukan target pasar wisata spesifik, mengoptimalkan pendapatan, mengatur pengeluaran pemasaran, menentukan strategi berikutnya, dan merancang layanan serta produk yang sesuai untuk setiap segmennya.

Peneliti menggunakan metode *clustering K-Means* untuk meningkatkan akurasi pengelompokan atau segmentasi pengunjung. Selanjutnya, hasil analisis *clustering* menggunakan *sub-cluster* dibandingkan dengan tanpa *sub-cluster*. Kemudian,model *clustering* optimal diimplementasikan sebagai salah satu fitur segmentasi pada website Lejja menggunakan Flask dan Laravel untuk visualisasi analisis segmentasi pengunjung.

Dalam proses *clustering* menggunakan *K-Means*, peneliti melakukan *data cleaning*, *feature engineering*, *data transformation*, dan *data reduction* guna meningkatkan kinerja *K-Means*. Selain itu, peneliti melakukan uji *ANOVA* dan *post hoc* dalam menentukan potensi pembentukan *sub-cluster*. Peneliti juga melakukan perbandingan *clustering* dengan *sub-cluster* dan tanpa *sub-cluster*.

Penelitian ini menggunakan 1000 data dan 16 variabel. Hasil penelitian ini menunjukkan *clustering* dengan *sub-cluster* memiliki hasil *clustering* lebih baik daripada *clustering* tanpa *sub-cluster* dari performa *clustering*, uji ANOVA, dan interpretasi *clustering*. *Clustering* dengan *sub-cluster* mencapai akurasi *silhouette score* 47% dengan membagi data menjadi 3 *cluster*. Selanjutnya, peneliti membentuk *sub-cluster* pada setiap *cluster* karena terdapat variasi tinggi pada beberapa fitur dalam setiap *cluster*. Pada *cluster* 0, rata-rata *WTP* adalah Rp. 6.727,05, dengan *sub-cluster* 0 sebesar Rp. 5.780,63 dan *sub-cluster* 1 sebesar Rp. 8.214,29. Pada *cluster* 1, rata-rata *WTP* adalah Rp. 12.307,69, dengan *sub-cluster* 0 sebesar Rp. 11.780,30, *sub-cluster* 1 sebesar Rp. 14.578,95, dan *sub-cluster* 2 sebesar Rp. 10.203,49. Sedangkan pada *cluster* 2, rata-rata *WTP* adalah Rp. 6.098,82, dengan *sub-cluster* 0 sebesar Rp. 7.240,74, *sub-cluster* 1 sebesar Rp. 5.286,89, dan *sub-cluster* 2 sebesar Rp. 5.426,83. Model *K-Means* optimal dideploy ke website Lejja.

Kata Kunci: *K-Means*, *WTP*, Segmentasi

ABSTRACT

WIRA DRANA WASISTHA. Segmentation of Tourist Visitors Based on Website-Based *Willingness to Pay* Using K*Mean*s Model (Case Study of Lejja Nature Tourism Park) (supervised by Indrabayu and Elly Warni)

Lejja Nature Park (TWA) managers have conducted research on the Willingness to Pay (WTP) of visitors in conservation areas for the improvement and addition of basic public facilities, such as entrance gates, road networks, Community Empowerment Areas (CEA), toilets, mosques/mushrooms, parking lots, and gazebos/shelters. Managers need further research to analyze visitor segmentation based on WTP to determine specific tourism target markets, optimize revenue, manage marketing expenditures, determine the next strategy, and design appropriate services and products for each segment.

Researchers use the K-Means clustering method to improve the accuracy of visitor grouping or segmentation. Furthermore, the results of clustering analysis using sub-clusters are compared with those without sub-clusters. Then, the optimal clustering model is implemented as one of the segmentation features on the Lejja website using Flask and Laravel to visualize visitor segmentation analysis.

In the clustering process using K-Means, we performed data cleaning, feature engineering, data transformation, and data reduction to improve K-Means performance. In addition, researchers conducted ANOVA and post hoc tests in determining the potential for sub-cluster formation. We also compared clustering with sub-cluster and without sub-cluster.

This study used 1000 data and 16 variables. The results showed that clustering with sub-clusters has better clustering results than clustering without sub-clusters from clustering performance, ANOVA test, and clustering interpretation. Clustering with sub-clusters achieved 47% silhouette score accuracy by dividing the data into 3 clusters. Furthermore, researchers formed sub-clusters in each cluster because there was high variation in some features within each cluster. In cluster 0, the average WTP is Rp. 6,727.05, with sub-cluster 0 of Rp. 5,780.63 and sub-cluster 1 of Rp. 8,214.29. In cluster 1, the average WTP is Rp. 12,307.69, with sub-cluster 0 amounting to Rp. 11,780.30, sub-cluster 1 amounting to Rp. 14,578.95, and sub-cluster 2 amounting to Rp. 10,203.49. While in cluster 2, the average WTP is Rp. 6,098.82, with sub-cluster 0 amounting to Rp. 7,240.74, sub-cluster 1 amounting to Rp. 5,286.89, and sub-cluster 2 amounting to Rp. 5,426.83. The optimal K-Means model was deployed to the Lejja website.

Keywords: *K-Means, WTP*, Segmentation

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
WTP	*Willingness to Pay*
CVM	*Contingent Valuation*
MSE	*Mean Square Error*
R2	Koefisien Determinasi
BKSDA	Balai Konservasi Sumber Daya Alam
E*WTP*	Rataan *WTP*(Rp)
W_i	Nilai *WTP* ke-I (Rp)
n	Jumlah responden (orang)
i	Urutan responden (i=1,2,…..,n)
PCA	*Principal Component Analysis*
x_i **dan** y_i	Koordinat titik *x* dan *y*
n	Jumlah dimensi
a	Jarak rata-rata sampel ke semua titik lainnya dalam *cluster* yang sama
b	Jarak rata-rata sampel ke semua titik dalam *cluster* terdekat yang bukan *cluster*nya
σ_i	Rata-rata jarak semua elemen dalam *cluster i* ke pusat *cluster* c_i,
$d(c_i,c_i)$	Jarak antara pusat *cluster i* dan *j*.
F	Koefisien *ANOVA*
MSB	Rata-rata jumlah kuadrat antar kelompok
HSD	Beda Nyata Terkecil (*Honestly Significant Difference*),
q	Nilai kritis dari tabel rentang yang disederhanakan,

DAFTAR LAMPIRAN

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir dengan judul "Segmentasi Pengunjung Wisata Berdasarkan *Willingness to Pay* Berbasis Website Menggunakan Model *KMeans* (Studi Kasus Taman Wisata Alam Lejja)". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini, penulis ingin menyampaikan terima kasih setulus-tulusnya kepada:

1. Allah SWT atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Kedua orangtua penulis, Bapak Alm. Syarifuddin dan Ibu Hj. Sumarni, beserta keluarga yang selalu mendukung penulis dalam menempuh pendidikannya, mendoakan demi kelancaran urusan perkuliahan, serta memberi semangat saat penulis mengerjakan skripsi. Penulis tidak akan sampai di titik ini tanpa doa dan restu kedua orangtua, sehingga penulis mengucapkan banyak terima kasih atas jasa dan kerja keras mereka.
3. Bapak Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM. ASEAN. Eng selaku pembimbing I dan Kepala Departemen Teknik Informatika, serta Ibu Elly Warni S.T,. M.T. selaku pembimbing II, yang telah menjadi mentor dalam berbagai hal, menyediakan waktu, tenaga, pikiran, dan perhatian luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
4. Segenap Dosen dan Staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu penulis selama masa perkuliahan.
5. Pak Sofyan, Kak Yusuf, Dea, Debi, Accung, Rahma, Ical, dan Anlin yang telah membantu dalam penelitian dan penyusunan tugas akhir.
6. Teman, rekan, dan keluarga Lab AIMP, Animasi, dan Ubicon yang telah berkontribusi dalam pengembangan diri peneliti.

Penulis berharap semoga Tuhan membalas segala kebaikan yang telah diterima dari berbagai pihak yang telah membantu mempermudah penulis dalam mengerjakan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu, penulis mengharapkan segala bentuk saran dan masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberikan pengetahuan dan manfaat bagi penulis dan pembaca.

Gowa, 5 Mei 2024

Wira Drana Wasistha

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pariwisata merupakan salah satu industri yang dinamis dan tercepat dalam mendorong pertumbuhan ekonomi global, khususnya di Indonesia. Indonesia mengalami kenaikan 12 peringkat, menjadi peringkat ke-32 dari total 117 negara dalam *Travel and Tourism Competitiveness Index* (*TTCI*) (Kemenparekraf, 2023). Dampak ekonomi pariwisata Indonesia diprediksi mencapai Rp. 2.000 triliun pada tahun 2028 berdasarkan riset yang dilakukan oleh Statista (Statista, 2024). Fakta ini meningkatkan persaingan antar perusahaan dalam mengembangkan strategi untuk meningkatkan pangsa pasar. Salah satu strategi yang digunakan adalah menganalisis profil dan kebiasaan pengunjung wisata dengan menggunakan *clustering* (Yildirim dkk., 2022).

Perusahaan Taman Wisata Alam Lejja (TWA) yang berlokasi di Desa Bulu'E, Kabupaten Soppeng, Sulawesi Selatan, Indonesia, telah melakukan penelitian terhadap kesediaan membayar (*Willingness to Pay/WTP*) pengunjung untuk menghitung kesediaan membayar pengunjung saat berada di kawasan konservasi untuk perbaikan dan penambahan fasilitas umum dasar, seperti gerbang masuk, jaringan jalan, *Community Empowerment Area (CEA)*, toilet, masjid/mushola, tempat parkir, dan gazebo/*shelter* (Perusda, 2022). Kesediaan untuk membayar atau *Willingness to Pay* (*WTP*) dalam konteks pariwisata mengacu pada jumlah uang yang bersedia dibelanjakan oleh wisatawan untuk meningkatkan keberlanjutan, memperkaya pengalaman wisata, atau mengakses fasilitas atau layanan tertentu di suatu destinasi. *WTP* yang tidak akurat dapat menyebabkan pendapatan tidak optimal, pengalokasian sumber daya yang tidak efisien, perencanaan dan pengembangan yang tidak efektif, dampak lingkungan yang tidak terkendali, ketidakpuasan pelanggan, kesulitan dalam pembiayaan pelestarian dan peningkatan, serta persaingan pasar yang tidak efektif (Durán-Román dkk., 2021).

Dari penelitian sebelumnya, didapatkan *WTP* sebesar Rp. 8.000 menggunakan metode *Contingent Valuation Method (CVM)*. Sehingga rekomendasi harga tiket adalah Rp. 20.000 (harga tiket saat ini) ditambah Rp. 8.000

(*WTP*), menjadi harga tiket sebesar Rp. 28.000. Perusahaan mengalami kesulitan dalam mengoptimalkan pendapatan, menentukan strategi berikutnya, serta memberikan produk dan layanan yang sesuai dengan pengunjung Lejja. Oleh karena itu, diperlukan analisis terhadap profil dan kebiasaan pengunjung untuk memberikan pendapatan optimal, mengoptimalkan pengeluaran pemasaran, dan menentukan strategi berikutnya, sehingga perusahaan dapat bersaing dengan kompetitornya (Perusda, 2022). Selain itu, perusahaan dapat mendesain layanan dan produk yang sesuai pada setiap segmen pengunjung (Farina Srihadi dkk., 2016). Hal tersebut dapat dilakukan dengan mengelompokkan pengunjung ke dalam beberapa *cluster* menggunakan *data mining* (Yildirim dkk., 2022).

Peneliti melakukan perbandingan penelitian terkait dalam melakukan *clustering* pengunjung untuk memberikan pertimbangan dalam penentuan metode *clustering*. Dalam penelitian "*A Case Study: Unsupervised Approach for Tourist Profile Analysis by K-Means Clustering in Turkey*," diperoleh hasil evaluasi silhouette score sebesar 32% dengan algoritma *K-Means* (Yildirim dkk., 2022). Selanjutnya, dalam penelitian "*Application of Clustering Algorithms on Tourism Industry*" mengaplikasikan algoritma *K-Means* dalam industri pariwisata (Tiwari & Tripathi, 2023). Perbandingan kinerja algoritma *K-Means* dengan algoritma lain, seperti *Support Vector Machine (SVM), K-Medoids, Nearest Neighbor,* dan *Naïve Bayesian* dalam melakukan *clustering* menunjukkan bahwa *K-Means* memiliki kinerja yang lebih baik, karena memiliki akurasi yang lebih tinggi dibandingkan algoritma lain (Jauhari dkk., 2022).

Maka dari itu, penelitian ini mengembangkan penelitian sebelumnya dengan menggunakan metode *K-Means* untuk melakukan *clustering*. Hasil dari pemanfaatan metode *K-Means* ini dapat menjadi solusi dalam mengelompokkan pengunjung berdasarkan *WTP*, seperti jenis kelamin, status, umur, lama pendidikan formal, pendidikan, jumlah tanggungan keluarga, pendapatan per bulan, lama mengetahui TWA, total kunjungan, kunjungan setahun terakhir, biaya wisata, waktu tempuh perjalanan, motivasi, kepuasan terhadap lingkungan dan fasilitas TWA, kepuasan terhadap kegiatan rekreasi, waktu survei, dan *WTP*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana melakukan segmentasi pengunjung berdasarkan *WTP* dengan menggunakan metode *clustering K-Means* serta hasil optimal yang diperoleh?
2. Bagaimana melakukan perbandingan analisis hasil *clustering* dengan menggunakan *sub-cluster* dan tanpa *sub-cluster*?
3. Bagaimana cara mengembangkan fitur analisis segmentasi pengunjung berdasarkan *WTP* pada situs web Lejja?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membuat segmentasi pengunjung berdasarkan *WTP* menggunakan metode *clustering K-Means* dan mencapai hasil optimal.
2. Membuat perbandingan analisis hasil *clustering* dengan menggunakan *sub-cluster* dan tanpa *sub-cluster*.
3. Mengembangkan fitur segmentasi pengunjung berdasarkan *WTP* pada situs web Lejja.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat membantu:

1. Memberikan wawasan tentang karakteristik pengunjung dari setiap segmen dengan tingkat akurasi yang tinggi.
2. Memberikan wawasan perbandingan hasil *clustering* dengan menggunakan *sub-cluster* dan tanpa *sub-cluster*.
3. Menampilkan visualisasi hasil segmentasi pengunjung secara langsung di dalam website.

1.5 Ruang Lingkup

Ruang lingkup dari penelitian ini adalah:

1. Penelitian dilakukan di objek wisata Lejja.
2. Rekomendasi sistem berupa fitur analisis pada website untuk segmentasi pengunjung berdasarkan *WTP*.
3. Penerapan metode *clustering* menggunakan *K-Means*.
4. Parameter yang digunakan dalam proses *clustering* pengunjung mencakup jenis kelamin, status, umur, lama pendidikan formal, pendidikan, jumlah tanggungan keluarga, pendapatan per bulan, lama mengetahui TWA, total kunjungan, kunjungan setahun terakhir, biaya wisata, waktu tempuh perjalanan, motivasi, kepuasan terhadap lingkungan dan fasilitas TWA, kepuasan terhadap kegiatan rekreasi, waktu survei, dan *WTP*.

BAB II
TINJAUAN PUSTAKA

2.1 Taman Wisata Alam Lejja

Menurut Undang-Undang Republik Indonesia Nomor 5 Tahun 1990 tentang Konservasi Sumber Daya Alam Hayati dan Ekosistemnya, Taman Wisata Alam (TWA) adalah kawasan pelestarian alam yang terutama dimanfaatkan untuk pariwisata dan rekreasi alam. Undang-undang republik indonesia nomor 5 tahun 1990 tentang konservasi sumber daya alam hayati dan ekosistemnya menyebutkan taman wisata alam berfungsi untuk kepentingan penelitian, ilmu pengetahuan dan pendidikan, serta menunjang budidaya dan wisata alam. Menurut Undang-Undang Republik Indonesia 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.

Nilai ekonomi wisata merupakan ukuran moneter dari total manfaat yang diperoleh dari suatu sumber daya wisata. Nilai ini tidak hanya mencakup pendapatan langsung dari aktivitas wisata seperti akomodasi, transportasi, makanan, dan tiket masuk atraksi, tetapi juga manfaat tidak langsung dan non-pasar seperti nilai konservasi, nilai estetika, dan kontribusi terhadap kesejahteraan sosial. Untuk mengukur nilai ekonomi wisata, digunakan pendekatan perhitungan *WTP* (*Willingness to Pay*) (Perusda, 2022).

TWA Lejja adalah pemandian air panas yang terletak di Kabupaten Soppeng, Sulawesi Selatan (Sulsel), yang dikelola oleh Perusda Soppeng atau PT Lamatesso Matappa. Sumber air panas Lejja ini berasal dari gunung api yang kini sudah tidak aktif. Lejja berada dalam kawasan Balai Konservasi Sumber Daya Alam (BKSDA) Provinsi Sulawesi Selatan. Di dalam kawasan ini terdapat 16 satwa liar yang dilindungi, termasuk 10 ekor kringking bukit (*prionitas platury*), dan satu ekor prikici dora (*tricholusus otaney*). Lokasi Pemandian Air Panas Lejja terletak di kawasan hutan lindung di Desa Bulue, Kecamatan Marioriawa, Kabupaten Soppeng (Perusda, 2022).

Pengelola Taman Wisata Alam Lejja yang berlokasi di Desa Bulu'E, Kabupaten Soppeng, Sulawesi Selatan, Indonesia, telah melakukan penelitian *Willingness to Pay* (*WTP*) terhadap pengunjung untuk menghitung kesediaan mereka membayar saat berada di kawasan konservasi guna perbaikan dan penambahan fasilitas umum dasar, seperti gerbang masuk dan jaringan jalan, area pemberdayaan masyarakat (*Community Empowerment Area/CEA*), toilet, masjid/mushola, tempat parkir, serta gazebo/*shelter* . Hasil penelitian menunjukkan bahwa *WTP* pengunjung sebesar Rp. 8.000. Oleh karena itu, rekomendasi untuk harga tiket adalah Rp. 20.000 (harga tiket saat ini) ditambah Rp. 8.000 (*WTP*), sehingga totalnya menjadi Rp. 28.000 (Perusda, 2022).

2.2 *Willingness to Pay Tourist*

Kesediaan untuk membayar atau *Willingness to Pay* (*WTP*) dalam konteks pariwisata merujuk pada jumlah uang yang siap dibelanjakan oleh wisatawan untuk meningkatkan keberlanjutan, memperkaya pengalaman wisata, atau mendapatkan akses ke fasilitas atau layanan tertentu di suatu destinasi (Durán-Román dkk., 2021). Rata-rata *WTP* dari keseluruhan data dihitung menggunakan persamaan (1) berikut (Perusda, 2022).

$$EWTP = \frac{\sum_{i-1}^{n} W_i}{n} \quad (1)$$

dimana,

EWTP = rataan *WTP*(Rp),
W_i = nilai *WTP* ke-I (Rp),
n = jumlah responden (orang),
i = urutan responden (i=1,2,…..,n).

2.3 Segmentasi Pengunjung

Pengelompokan atau segmentasi merupakan salah satu konsep terpenting dalam pemasaran (Dolnicar, 2020). Tujuan dari segmentasi adalah untuk membagi populasi menjadi kelompok-kelompok yang dapat dibedakan satu sama lain, sehingga bisnis dapat memposisikan dirinya secara unik dan memahami dengan lebih efektif dan efisien tentang kebutuhan, karakteristik, atau perilaku konsumen yang berbeda yang memerlukan produk dan pemasaran yang berbeda. Dalam

konteks segmentasi pengunjung wisata, perusahaan perlu menganalisis profil dan kebiasaan pengunjung yang dapat memberikan pendapatan optimal, mengoptimalkan pengeluaran pemasaran, dan merencanakan strategi selanjutnya, sehingga perusahaan dapat bersaing dengan kompetitornya (Perusda, 2022). Selain itu, perusahaan juga dapat merancang layanan dan produk yang sesuai untuk setiap segmennya (Farina Srihadi dkk., 2016). Segmentasi pengunjung wisata umumnya dilakukan dengan menggunakan pendekatan *cluster*, seperti yang dilakukan dalam penelitian segmentasi pasar pariwisata perkotaan yang dilakukan oleh (Carvache-Franco dkk., 2023).

2.3.1 Tipe Segmentasi Pengunjung

Tipe segmentasi pengunjung sebagai berikut (Nurjannah dkk., 2019).

a. Segmentasi Geografi

   Segmentasi geografi mengelompokkan pasar berdasarkan unit geografi seperti negara, wilayah regional, kabupaten, kota, atau tempat tinggal.

b. Segmentasi Demografi

   Segmentasi ini mengelompokkan pasar berdasarkan berbagai variabel seperti usia, jenis kelamin, tingkat pendapatan, pekerjaan, tingkat pendidikan, dan sebagainya.

c. Segmentasi Psikografi

   Segmentasi ini mengelompokkan konsumen berdasarkan kelas sosial, gaya hidup, minat dan hobi, serta penampilan.

d. Segmentasi Perilaku

   Pada segmentasi ini, konsumen dikelompokkan berdasarkan frekuensi perjalanan, tujuan kunjungan, pengeluaran, tipe transportasi, dan lainnya.

2.4 *Preprocessing Data*

Preprocessing data atau pra pemrosesan dalam konteks *data science* dan *machine learning* adalah serangkaian prosedur yang dilakukan pada data sebelum diterapkan ke algoritma *machine learning*. Tujuan dari *Preprocessing* adalah untuk membuat data menjadi lebih sesuai dan efektif untuk proses pembelajaran. Ini adalah langkah kritis dalam *pipeline machine learning* karena kualitas dan format

data secara langsung mempengaruhi kinerja dan keakuratan model. Berikut merupakan proses *preprocessing* sebagai berikut (García dkk., 2016).

1. *Data Cleaning*

   Pembersihan data atau *data cleaning* bertujuan untuk menghilangkan atau memperbaiki data yang hilang atau tidak konsisten, termasuk pengisian nilai yang hilang, menghapus baris atau kolom, atau memperbaiki format data yang salah (Buuren, S, 2018). Proses dalam *data cleaning* sebagai berikut.

   a. *Cleaning variable names*

      Proses ini melibatkan penggantian nama kolom menjadi label yang lebih mudah dikenali, sehingga data lebih mudah dibaca dan dipahami (rdrr, 2024).

   b. *Missing value handling*

      Salah satu cara pengisian nilai yang hilang yaitu menggunakan metode imputasi. Tujuan dari imputasi adalah untuk menghasilkan dataset yang lengkap sehingga analisis statistik atau model *machine learning* dapat dijalankan dengan lebih efektif. Salah satunya metode imputasi median yaitu menggunakan median dari data yang tersedia untuk mengisi nilai yang hilang. Ini sering digunakan ketika distribusi data tidak simetris atau memiliki outlier (Buuren, S, 2018).

2. *Feature Engineering*

   Feature engineering atau rekayasa fitur adalah proses memanipulasi, menambah, menghapus, menggabungkan, atau mengubah fitur dalam dataset untuk meningkatkan pelatihan model *machine learning*, yang mengarah pada kinerja yang lebih baik dan akurasi yang lebih tinggi (aws, 2024). Proses yang termasuk *feature engineering* sebagai berikut.

   a. *Feature creation*

      Feature creation melibatkan proses pembuatan fitur baru dari data yang ada untuk meningkatkan kekuatan prediksi model pembelajaran mesin. *Feature creation* merupakan aspek penting dari pra pemrosesan data dan pengembangan model, yang bertujuan untuk menyediakan kumpulan data yang lebih komprehensif dan informatif untuk melatih algoritma *machine learning* (aws, 2024).

b. *Binning*

Binning adalah proses mengubah variabel numerik menjadi variabel kategorik dengan mengelompokkan data numerik kontinu ke dalam interval diskrit (aws, 2024).

3. *Data Transformation*

Data transformation atau transformasi data mengacu pada proses mengubah dan memanipulasi data dari satu format ke format lainnya, untuk meningkatkan kualitas dan kegunaannya (Pratt & Bernstein, 2024). Berikut proses *data transformation* sebagai berikut.

a. *Label encoding*

Label encoding bertujuan untuk mengubah variabel kategorikal menjadi format numerik menggunakan teknik seperti *one-hot encoding* atau *label encoding*. *Label encoding* lebih baik digunakan ketika variabel kategorikal memiliki urutan alami (misalnya peringkat seperti rendah, sedang, tinggi). Sedangkan *one-hot encoding* lebih sesuai untuk variabel kategorikal di mana tidak ada urutan atau hierarki, dan di mana jumlah kategori tidak terlalu besar sehingga tidak menimbulkan masalah dimensi tinggi (Rodríguez dkk., 2018).

Dalam beberapa kasus langkah awal dalam *label encoding* menggunakan *dictionary mapping*. *Dictionary mapping* adalah proses membuat peta atau kamus di mana setiap kunci unik (dalam kasus ini, setiap label kategori) dikaitkan dengan nilai tertentu (dalam konteks ini, nilai numerik). *Dictionary mapping* berguna ketika dataset memiliki banyak kategori atau jika ada kebutuhan untuk mempertahankan konsistensi *mapping* dalam dataset yang berbeda atau sepanjang waktu (Rodríguez dkk., 2018).

b. *Data type conversion*

Data type conversion atau konversi tipe data dalam transformasi data mengacu pada proses penerjemahan data dari satu format ke format lain untuk memungkinkan eksekusi yang tepat dalam file, aplikasi, atau basis data. Ini melibatkan konversi data dalam kolom input ke tipe data yang berbeda dan

kemudian menyalinnya ke kolom *output* baru, sebuah langkah umum dalam pembersihan data untuk membuat data lebih sesuai untuk analisis, visualisasi, dan interpretasi (Pratt & Bernstein, 2024).

c. *Data intersection*

Data intersection dalam *feature engineering* diperlukan untuk menggabungkan dua set data berdasarkan fitur bersama. *Intersection* akan memastikan bahwa hanya entri yang memiliki fitur bersama yang akan digabungkan, menjaga integritas data (Pratt & Bernstein, 2024).

d. *Data concatenation*

Data concatenation atau penggabungan data mengacu pada proses menggabungkan atau menghubungkan data dari sumber yang berbeda atau dalam set data yang sama. Hal ini dapat melibatkan penumpukan kumpulan data di atas satu sama lain untuk membuat satu kumpulan data terpadu (IBM, 2015).

e. *Normalization*

Normalization bertujuan dalam penskalaan data numerik dalam rentang tertentu tanpa mengubah bentuk distribusi data, sehingga mengubah skala nilai fitur dan fitur dengan skala besar tidak mendominasi saat dilakukan penghitungan matematis, khususnya dalam model *machine learning*. Metode yang digunakan yaitu *StandardScaler*. *StandardScaler* berguna karena data memiliki distribusi normal dan algoritma *K-Means* bergantung pada asumsi distribusi normal. *StandardScaler* tidak memiliki rentang tetap seperti beberapa *scaler*. Sebaliknya, *StandardScaler* mengubah data sehingga distribusinya memiliki *mean* (rata-rata) 0 dan standar deviasi 1 (Mishra dkk., 2020).

4. *Data Reduction*

Reduksi data adalah proses mengurangi ukuran atau kompleksitas set data dengan tetap mempertahankan informasi yang paling penting. Tujuan utama dari reduksi data adalah untuk mengoptimalkan ruang penyimpanan, meningkatkan kecepatan pemrosesan data, dan meningkatkan efisiensi komputasi. Namun, reduksi data berpotensi mengakibatkan hilangnya informasi. Salah satu metode reduksi yaitu *Principal Component Analysis* (*PCA*) (Yu & Posey, 2024).

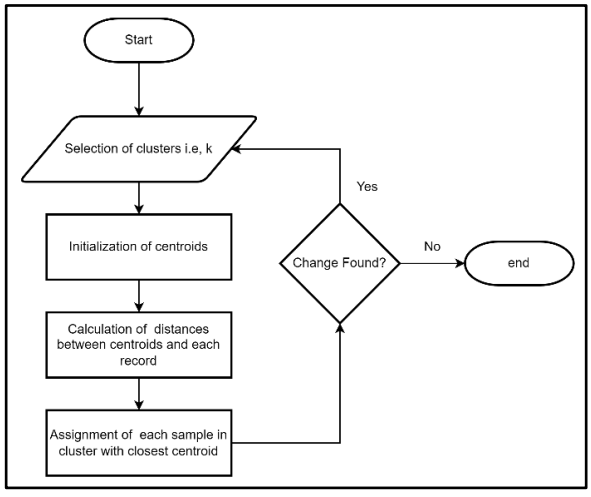
2.5 *Clustering*

Clustering adalah teknik yang digunakan dalam analisis data untuk mengelompokkan objek-objek yang serupa menjadi kelompok atau *cluster*. Tujuan utamanya adalah untuk memastikan bahwa data dalam setiap *cluster* serupa satu sama lain dan berbeda dengan data di *cluster* lain. Berikut merupakan jenis-jenis *clustering* (Patel dkk., 2015).

a. *Hierarchical clustering*: Membentuk hirarki *cluster*, bisa *divisive* (*top-down*) atau *agglomerative* (*bottom-up*).
b. *Partitional clustering*: Membagi data menjadi beberapa *cluster* tanpa membentuk hirarki seperti metode *K-Means*.

2.6 *K-Means*

Algoritma *K-Means* adalah metode yang digunakan untuk mengelompokkan data berdasarkan pencarian iteratif untuk pusat *cluster*. Lokasi *cluster* setiap data dihitung dengan mencari jarak minimum dari setiap titik data ke pusat *cluster* (Jauhari dkk., 2022). Pengelompokan informasi dengan menggunakan metode *K-Means*, secara umum dapat dilakukan dengan algoritma seperti *Flowchart* pada Gambar 1.



Gambar 1 *Flowchart K-Means*

Sumber: (Zhang dkk., 2021)

Berdasarkan *Flowchart* pada Gambar 1 dijelaskan sebagai berikut (Jauhari dkk., 2022).

a. Tentukan jumlah *cluster* awal.

   Metode yang digunakan dalam penentuan jumlah *cluster* pada penelitian ini yaitu *elbow method*. Tujuannya adalah untuk menentukan nilai k (jumlah *cluster*) yang sesuai dengan mengoptimalkan kriteria tertentu. Tentukan *centroid* sesuai dengan jumlah *cluster*.

b. Hitung jarak antara data dengan pusat *cluster* sesuai dengan *centroid* menggunakan *euclidean distance*.
c. Mengalokasikan data ke dalam *cluster* sesuai dengan *centroid* baru menggunakan perhitungan jarak terdekat.
d. Kembali ke langkah 2 jika masih ada data yang berpindah *cluster* atau terjadi perubahan nilai *centroid* di atas nilai *threshold* yang ditentukan atau jika perubahan nilai fungsi objektif yang digunakan di atas nilai ambang batas yang digunakan.

2.7 *Principle Component Analysis (PCA)*

Principal Component Analysis (*PCA*) adalah teknik reduksi dimensi linier yang digunakan dalam analisis data dan *machine learning*. Teknik ini melibatkan transformasi data ke dalam sistem koordinat baru untuk mengungkapkan struktur data yang mendasarinya. Rumus untuk *PCA* melibatkan beberapa langkah, termasuk perhitungan *covariance matrix*, *eigenvectors* dan *eigenvalues*, dan transformasi data asli ke dalam sistem koordinat yang baru. Berikut ini adalah langkah-langkah utama yang terlibat dalam proses *PCA* (Aluja-Banet dkk., 2018).

a. *Covariance matrix* menghitung matriks kovarian dari data asli.
b. *Eigenvectors* dan *eigenvalues* menghitung vektor eigen dan nilai eigen dari matriks kovarians. Vektor eigen mewakili komponen utama, dan nilai eigen menunjukkan jumlah varians yang dijelaskan oleh setiap komponen utama.
c. Proyeksi bertujuan memproyeksikan data asli ke sistem koordinat baru yang ditentukan oleh vektor eigen untuk mendapatkan komponen utama.

2.8 *Elbow Method*

Elbow method adalah teknik yang digunakan dalam analisis *cluster* untuk menentukan jumlah *cluster* yang optimal dalam sebuah set data. Metode ini melibatkan pemetaan variasi yang dijelaskan sebagai fungsi dari jumlah *cluster* dan mengidentifikasi titik "siku", yang mewakili jumlah *cluster* yang optimal. Metode siku biasanya digunakan dalam pengelompokan *K-Means* untuk menemukan nilai terbaik dari K, jumlah *cluster*. Metode ini merupakan metode grafis yang membantu dalam mengidentifikasi jumlah *cluster* yang sesuai berdasarkan nilai *Within Cluster Sum of Square* (*WCSS*) (Syakur dkk., 2018).

2.9 *Within Cluster Sum of Squares (WCSS)*

Within Cluster Sum of Squares (*WCSS*) adalah ukuran total variasi dalam data yang dijelaskan oleh *cluster*. *WCSS* adalah jumlah total kuadrat jarak antara setiap anggota *cluster* dan pusatnya. Tujuan dari *K-Means* dan banyak algoritma *cluster* lainnya adalah untuk meminimalkan nilai *WCSS*, yang berarti mengoptimalkan penempatan *cluster* sehingga total jarak antara titik-titik dalam *cluster* ke pusat *cluster* (*centroid*) adalah yang terendah. *WCSS* adalah indikator seberapa baik atau erat data dalam *cluster*. Semakin rendah *WCSS*, semakin serupa atau lebih dekat data dalam satu *cluster*. Persamaan *WCSS* ditunjukkan seperti pada persamaan (2) (Kuraria dkk., 2018).

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

dimana,

k = jumlah *cluster*,

C_i = himpunan titik data dalam jumlah *cluster i*,

x = titik data individu dalam *cluster i*,

μ_i = *centroid cluster i*,

$||x - \mu_i||^2$ = kuadrat jarak *Euclidean* antara titik data x dan *centroid* μ_i.

2.10 *Euclidean Distances*

Euclidean Distances adalah metode yang paling umum digunakan untuk mengukur jarak antara dua titik dalam ruang *euclidean*, yaitu ruang yang digambarkan dalam geometri biasa. Ini adalah cara standar untuk mengukur "jarak

sebenarnya" antara dua titik, tidak peduli dimensi ruang tersebut. Jarak *euclidean* antara dua titik dalam ruang 2D atau 3D dapat dihitung dengan menggunakan *teorema pythagoras*. Rumus umum untuk menghitung jarak *euclidean* antara dua titik seperti pada persamaan (3) berikut (Jauhari dkk., 2022).

$$Dist\,(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

dimana,

x_i = koordinat titik x,

y_i = koordinat titik y,

n = jumlah dimensi.

2.11 *Silhouette Score*

Silhouette score adalah metode untuk menilai kualitas pengelompokan yang dilakukan oleh algoritma *clustering*. Nilai *Silhouette score* mengukur seberapa serupa sebuah objek dengan *cluster*nya sendiri dibandingkan dengan *cluster* lainnya. Skor ini berkisar antara -1 hingga 1, di mana nilai yang tinggi menandakan bahwa objek cocok dengan baik dengan *cluster*nya sendiri dan tidak cocok dengan *cluster* tetangga (Shahapure & Nicholas, 2020). Untuk setiap sampel, *Silhouette score score* dihitung seperti persamaan (4).

$$\mathcal{S} = \frac{(b-a)}{\max(a,b)} \quad (4)$$

dimana,

a = jarak rata-rata sampel ke semua titik lainnya dalam *cluster* yang sama,

b = jarak rata-rata sampel ke semua titik dalam *cluster* terdekat yang bukan *cluster*nya.

Untuk menilai suatu nilai *Silhouette Score Coefficient*, dapat dilihat pada Tabel 1 berikut.

Tabel 1 Interpretasi Nilai *Silhouette Score Coefficient*

Nilai *Silhouette Score Coefficient*	**Interpretasi**
0,71 – 1,00	*Cluster* yang kuat
0,51 – 0,70	*Cluster* telah baik atau sesuai
0,26 – 0,50	*Cluster* yang lemah

Nilai *Silhouette Score Coefficient*	Interpretasi
≤ 0,25	Tidak dapat dikatakan sebagai *Cluster*

Sumber: (Azuri & Pontoh, 2016)

2.11 *Davies-Bouldin Index* (DBI)

Davies-Bouldin Index (DBI) adalah metode lain untuk mengevaluasi kualitas pengelompokan. DBI merupakan rasio jarak *intra-cluster* dan *inter-cluster*. Indeks ini meminimalkan jarak *intra-cluster* dan memaksimalkan jarak antar *cluster*, *cluster* yang lebih terpisah dan lebih padat memberikan nilai DBI yang lebih rendah, menandakan pengelompokan yang lebih baik. Indeks *davies-bouldin* didefinisikan sebagai rata-rata *'similarity measure*' antara setiap *cluster* dengan *cluster* terdekatnya, di mana *similarity measure* ini adalah rasio antara sum intra-*cluster* dan antara *cluster* (Wei dkk., 2020). Perhitungan DBI berdasarkan persamaan (5) berikut.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right) \quad (5)$$

dimana,

σ_i = rata-rata jarak semua elemen dalam *cluster i* ke pusat *cluster* c_i,

$d(c_i,c_i)$ = jarak antara pusat *cluster i* dan *j*.

2.12 *Analysis of Variance* (*ANOVA*)

ANOVA, singkatan dari "*Analysis of Variance*," adalah metode statistik yang digunakan untuk membandingkan rata-rata antara dua atau lebih kelompok. Tujuannya adalah untuk mengetahui apakah ada perbedaan yang signifikan secara statistik antara rata-rata kelompok. Prinsip dasar *ANOVA* yaitu variasi antar kelompok yang menilai seberapa jauh rata-rata kelompok berbeda dari rata-rata keseluruhan dan variasi dalam kelompok yang menilai variabilitas pengamatan dalam setiap kelompok. Dalam pengujian *ANOVA* dilakukan menggunakan hipotesis sebagai berikut (Kim, 2017).

1. *Null hypothesis* (H0): Rata-rata semua kelompok adalah sama.

2. *Alternative hypothesis* (H1): Setidaknya satu rata-rata kelompok berbeda dari yang lain.

ANOVA memiliki beberapa jenis, salah satunya *One-Way ANOVA*. *One-Way ANOVA* digunakan ketika kita memiliki satu variabel independen kategorikal dengan tiga atau lebih level (grup) dan satu variabel dependen kontinu (Kim, 2017). Rumus utama yang digunakan dalam *One-Way ANOVA* dimulai menghitung *Sum of Square Between (SSB)* untuk mengukur variabilitas antara grup seperti pada persamaan (6) berikut (Sullivan, 2019).

$$SSB = \sum n_i(\bar{X}_i - \bar{X}_G)^2 \quad (6)$$

dimana,

n_i = jumlah sampel dalam grup i,

$\bar{X}_i$ = rata-rata sampel untuk grup I,

$\bar{X}_G$ = rata-rata keseluruhan.

Selanjutnya, dihitung variabilitas dalam grup atau *Sum of Square Within (SSW)* seperti pada persamaan (7) berikut (Sullivan, 2019).

$$SSW = \sum\left(\bar{X}_{ij} - \bar{X}_i\right)^2 \quad (7)$$

dimana,

$\bar{X}_{ij}$= nilai observasi j dalam grup i,

$\bar{X}_i$ = rata-rata sampel untuk grup I,

Selanjutnya, dihitung variabilitas total dalam data atau *Total Sum of Squares (SST)* seperti pada persamaan (8) berikut (Sullivan, 2019).

$$SST = SSB + SSW \quad (8)$$

Selanjutnya, dihitung *Degrees of Freedom Between (dfB)* seperti pada persamaan (9) berikut (Sullivan, 2019).

$$dfB = k - 1 \quad (9)$$

dimana,

k = Jumlah grup.

Selanjutnya, dihitung *Degrees of Freedom Within (dfW)* seperti pada persamaan (10) berikut (Sullivan, 2019).

$$dfW = N - k \quad (10)$$

dimana,

N = jumlah total observasi.

Selanjutnya, dihitung *Mean Square Between (MSB)* seperti pada persamaan (11) berikut (Sullivan, 2019).

$$MSB = \frac{SSB}{dfB} \tag{11}$$

Selanjutnya, dihitung *Mean Square Within (MSW)* seperti pada persamaan (12) berikut (Sullivan, 2019).

$$MSW = \frac{SSW}{dfW} \tag{12}$$

Selanjutnya, dihitung *F-Statitic* yang membandingkan varians antar kelompok dengan varians dalam kelompok seperti pada persamaan (13) berikut (Sullivan, 2019).

$$F = \frac{MSB}{MSW} \tag{13}$$

Selanjutnya nilai *p-value* diperoleh dengan membandingkan F-statistik yang dihitung dengan distribusi F. Ini dilakukan menggunakan tabel distribusi F atau software statistik. *P-value* menunjukkan probabilitas mendapatkan hasil observasi atau lebih ekstrim jika Hipotesis Nol benar. Jika *p-value* lebih kecil dari tingkat signifikansi yang ditetapkan (misalnya, 0.05), maka menolak Hipotesis Nol. Ini menunjukkan bahwa setidaknya satu kelompok memiliki rata-rata yang berbeda secara signifikan. Jika *p-value* lebih besar dari tingkat signifikansi, berarti memiliki cukup bukti untuk menolak Hipotesis Nol, yang berarti tidak dapat disimpulkan bahwa ada perbedaan yang signifikan antara rata-rata kelompok (Kim, 2017).

2.13 Uji *Post Hoc* menggunakan Uji *Tukey HSD*

Uji *post hoc*, yang berasal dari frasa Latin "setelah ini", adalah analisis statistik yang dilakukan setelah uji awal dilakukan. Uji ini digunakan untuk mengidentifikasi perbedaan spesifik antar kelompok, terutama ketika hipotesis nol dari uji *omnibus*, seperti *ANOVA*, ditolak. Uji *post hoc* juga dikenal sebagai uji perbandingan berganda dan digunakan untuk menentukan rata-rata kelompok mana yang secara signifikan berbeda satu sama lain. Uji *post hoc* termasuk salah satunya uji *HSD Tukey* dan lainnya. Tes-tes ini penting untuk mengeksplorasi perbedaan antara beberapa rata-rata kelompok sambil mengendalikan tingkat kesalahan eksperimen (Patel dkk., 2015).

Uji *Tukey's Honestly Significant Difference (HSD)* adalah uji *post hoc* yang biasa digunakan untuk menilai signifikansi perbedaan antara pasangan rata-rata kelompok. Uji ini membandingkan semua pasangan rata-rata yang mungkin dan didasarkan pada distribusi rentang yang disederhanakan. Uji *HSD* menggunakan berbagai fitur seperti perbedaan rata-rata (Perb Rataan), *p-value* terkoreksi (*p-adj*), serta batas bawah dan atas dalam analisis statistik memberikan pemahaman yang lebih komprehensif dan nuansa yang lebih dalam terkait hasil penelitian. *Tukey's HSD*, perbedaan rata-rata, *p-adj*, dan interval kepercayaan semuanya berkontribusi pada pengambilan keputusan tentang signifikansi statistik. *Tukey's HSD* dan *p-adj* langsung menangani signifikansi statistik perbandingan, sementara perbedaan rata-rata dan interval kepercayaan memberikan informasi tentang besarnya perbedaan dan keakuratan estimasi tersebut. Menggunakan semua informasi ini bersama-sama memungkinkan peneliti untuk membuat interpretasi yang lebih kaya dan lebih informasi tentang data peneliti (Abdi & Williams, 2010).

Perb Rataan memberikan informasi mengenai besarnya perbedaan antara grup. Ini membantu dalam memahami signifikansi praktis dari temuan tersebut, bukan hanya signifikansi statistiknya. Misalnya, sebuah perbedaan yang signifikan secara statistik mungkin tidak berarti banyak dalam praktik jika perbedaannya sangat kecil seperti ditunjukkan pada persamaan (14) berikut (Howell, 2010) .

$$Perb\ Rataan = \bar{x}_2 - \bar{x}_1 \tag{14}$$

dimana,

$\bar{x}_1$ = Rata-rata sampel untuk kelompok pertama,

$\bar{x}_2$ = Rata-rata sampel untuk kelompok kedua.

P-value Terkoreksi (*p-adj*) mengatasi masalah pengujian multipel, yang bisa meningkatkan risiko kesalahan tipe I (kesalahan positif palsu). Ini memberikan ukuran yang lebih konservatif dan akurat mengenai signifikansi statistik dari perbandingan, memastikan bahwa kesimpulan yang diambil dari data lebih dapat diandalkan. *P-adj* tidak dihitung melalui rumus yang sederhana seperti yang ditemukan dalam perhitungan statistik lainnya. Sebaliknya, *p-value* dalam uji *Tukey HSD* dihasilkan dari perbandingan distribusi probabilitas khusus yang terkait dengan perbedaan rata-rata antar pasangan grup. Prinsip bahwa perbedaan antara semua pasangan grup harus dianalisis dalam satu langkah, untuk mengontrol

keseluruhan tingkat kesalahan tipe I pada level yang diinginkan (biasanya α = 0.05). *P-adj* diwakili oleh probabilitas bahwa perbandingan spesifik antara grup akan menghasilkan nilai q yang besar atau lebih, jika hipotesis nol benar (tidak ada perbedaan). Nilai q yang dihitung kemudian dibandingkan dengan tabel distribusi q *Tukey*, yang menyediakan nilai kritis berdasarkan jumlah grup yang dibandingkan dan jumlah total observasi. Jika nilai q yang dihitung melebihi nilai kritis tersebut, maka *p-adj* untuk perbandingan tersebut akan menunjukkan signifikansi statistik. Rumus dari q ditunjukkan pada persamaan (15) berikut (Abdi & Williams, 2010) (Howell, 2010).

$$q = \frac{Perb\ Rataan\ Terbesar}{Estimasi\ Standar\ Kesalahan} \quad (15)$$

dimana,

q = Nilai statistik yang dihitung, yang kemudian dibandingkan dengan nilai kritis dari distribusi q *Tukey* untuk menentukan signifikansi.

MSE = *Mean* Square Error dari ANOVA,

n = Ukuran sampel untuk setiap grup.

Batas Bawah dan Atas merupakan Interval kepercayaan (yang diwakili oleh batas bawah dan atas) memberikan rentang nilai di mana perbedaan sebenarnya antara kelompok berada, dengan tingkat kepercayaan tertentu (biasanya 95%). Ini memberikan gambaran tentang ketidakpastian sekitar estimasi perbedaan rata-rata dan adalah penting untuk mengukur reliabilitas dari temuan tersebut. Perhitungan Batas Bawah atau Atas dimulai dengan menghitung estimasi standar kesalahan seperti pada persamaan (16) berikut (Abdi & Williams, 2010) (Howell, 2010).

$$Estimasi\ Standar\ Kesalahan = \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad (16)$$

dimana:

n_i dan n_j = Ukuran sampel dari dua kelompok yang dibandingkan,

Selanjutnya, perhitungan Batas Bawah atau Atas menggunakan persamaan (17) berikut (Howell, 2010).

$$Batas\ Bawah\ atau\ Atas = Perb\ Rataan\ \pm (q_{\propto}\ x\ Estimasi\ Standar\ Kesalahan) \quad (17)$$

dimana,

Simbol $\pm$ = Menunjukkan bahwa untuk mendapatkan batas bawah, dikurangkan nilai setelahnya, dan untuk batas atas, ditambahkan nilai setelahnya.

$q_{\propto}$ = Nilai kritis dari distribusi q *Tukey* untuk level signifikansi α yang diinginkan (misalnya, 0.05) berdasarkan derajat kebebasan total dan jumlah perbandingan yang dilakukan,