

Penerapan Algoritma *K-Modes Clustering* dengan Validasi *Davies Bouldin Index* pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta

Alyeska Astri Az-zahra

Politeknik Statistika STIS, 211709536@stis.ac.id

Almira Fajriyati Marsaoly

Politeknik Statistika STIS, 211709529@stis.ac.id

Intan Putri Lestyani

Politeknik Statistika STIS, 211709755@stis.ac.id

Roghibah Salsabila

Politeknik Statistika STIS, 211709986@stis.ac.id

Wa Ode Zuhayeni Madjida, S.ST, M.T.

Politeknik Statistika STIS, zuhayeni@stis.ac.id

ABSTRAK, *Kemajuan teknologi informasi dan komunikasi serta keberadaan internet yang semakin meluas di tengah masyarakat sehingga memunculkan kebiasaan-kebiasaan baru. Salah satunya adalah kebiasaan melakukan transaksi jual beli online. Kebiasaan baru tersebut menuntut para pemilik usaha untuk dapat menyesuaikan diri dengan perkembangan dunia transaksi saat ini. Penelitian ini bertujuan untuk mengetahui pengelompokan minat belanja online yang dilakukan masyarakat Daerah Istimewa Yogyakarta berdasarkan faktor jenis kelamin, umur, dan platform yang diminati dengan penerapan Data Mining. Dengan adanya pengelompokan ini diharapkan dapat memberikan pengetahuan bagi para pemilik online shop agar dapat mengetahui platform mana yang sering digunakan oleh masyarakat DIY dalam melakukan belanja online. Metode yang digunakan dalam mengelompokkan tingkat minat belanja online adalah *k-Modes Clustering* dengan nilai $k = 2,3,4,\dots,10$. *Davies-Boulden Index (DBI)* digunakan untuk menentukan jumlah kluster terbaik. Berdasarkan hasil analisis, diperoleh jumlah kluster terbaik yaitu $k=9$ dengan nilai *DBI* sebesar 1,3427. Kluster 5 merupakan kluster terbaik yang anggotanya sangat minat belanja melalui Marketplace dan Media Sosial. Marketplace yang diminati adalah Shopee, Bukalapak, dan Tokopedia, sedangkan Media Sosial yang diminati adalah Instagram, Facebook, dan Media Chatting. Kluster ini didominasi oleh laki-laki umur muda (15-24 tahun).*

Kata Kunci: *belanja online; mengelompokkan; k-modes; data mining*

1. PENDAHULUAN

Di era globalisasi saat ini, teknologi komunikasi dan informasi terus menerus berkembang menyesuaikan kebutuhan manumur dalam masyarakat. Berkembangnya teknologi

komunikasi dan informasi yang semakin pesat termasuk internet memberikan dampak yang besar dalam berbagai aspek kehidupan. Adanya perkembangan teknologi menjadikan internet tidak hanya digunakan untuk berkomunikasi saja namun dapat juga digunakan untuk melakukan transaksi jual beli.

Dengan adanya teknologi digital, dapat mendorong perekonomian suatu negara ke arah ekonomi digital. Sebenarnya era ekonomi digital sudah berlangsung sejak tahun 1980-an, dengan menggunakan *personal computer (PC)* dan internet sebagai teknologi kunci yang digunakan untuk efisiensi bisnis. Penggunaan teknologi seperti PC dan internet ini pun menjadi awal dari perkembangan *e-commerce*. Seiring dengan perkembangan teknologi, *era old digital economy* akhirnya memasuki *era new digital economy*, ditandai dengan adanya *mobile technology*, akses internet yang tidak terbatas, serta kehadiran teknologi *cloud* yang digunakan dalam proses ekonomi digital [22].

Berdasarkan data *Internet World Stats*, pengguna internet di Indonesia mengalami peningkatan dari tahun ketahun dimana pada tahun 2018 Indonesia menempati peringkat ke-5 negara dengan jumlah pengguna internet tertinggi di dunia yaitu sekitar 143 juta orang. Dengan perkembangan pengguna internet yang semakin pesat dapat menjadi sebuah potensi bisnis yang sangat menjanjikan. Jumlah pengguna internet yang terus meningkat dapat menjadi pasar yang potensial untuk dimasuki

para pebisnis dan aktualisasi belanja yang terjadi sangat cepat, nyaman, dan dapat dilakukan kapan saja dan dimana saja.

Perkembangan jumlah pengguna internet dapat mendorong terjadinya suatu potensi besar terciptanya usaha melalui media internet seperti *online shop* yang terdapat di berbagai platform, mulai dari media sosial hingga aplikasi chat. Dengan kemudahan akses yang ditawarkan oleh adanya usaha melalui internet, membuat masyarakat mulai merubah perilaku dalam belanja barang kebutuhan menjadi belanja secara *online* melalui handphone atau computer dengan memanfaatkan fasilitas internet.

Provinsi di Indonesia yang berpotensi untuk pengembangan usaha melalui internet adalah Daerah Istimewa Yogyakarta (DIY). Menurut data BPS, pada tahun 2018 DIY berada di posisi kedua terbesar di Indonesia setelah DKI Jakarta, dimana proporsi individu pengguna internet di DIY sebesar 55,45 persen. Hal ini juga diikuti dengan banyaknya jumlah usaha yang memanfaatkan internet dalam keperluan usahanya. Data statistik *e-commerce* BPS tahun 2019 menunjukkan bahwa persentase usaha yang melakukan *e-commerce* di DIY sebesar 27,88 persen, DIY memiliki persentase tertinggi se-Indonesia.

Berdasarkan hal tersebut peneliti ingin melakukan pengelompokan minat belanja *online* dilakukan berdasarkan faktor umur, jenis kelamin dan platform yang diminati dengan penerapan *Data Mining*. Dengan adanya pengelompokan diharapkan dapat memberikan pengetahuan bagi para pemilik *online shop* agar dapat mengetahui platform mana yang sering digunakan oleh masyarakat DIY dalam melakukan belanja *online*. Sehingga hal ini dapat bermanfaat dalam pengambilan keputusan berjualan *online* melalui platform apa, mengingat DIY merupakan provinsi yang berpotensi untuk pengembangan usaha melalui internet.

Metode statistika yang digunakan untuk pengelompokan minat belanja *online* adalah analisis kluster karena analisis kluster dapat mengelompokkan objek ke dalam kelompok yang relatif homogen [7]. Metode analisis kluster yang sering digunakan adalah analisis kluster *k-Means*, namun algoritma *k-Means* hanya bekerja

terbatas pada dataset yang atributnya bertipe numerik. Sedangkan untuk data bertipe kategorik menggunakan algoritma klustering *k-Modes* merupakan pengembangan dari algoritma *k-Means* [10]. Jumlah kluster dapat ditentukan sebanyak dua, tiga, dan seterusnya dimana setiap kluster memiliki karakteristik yang sama. Jumlah kluster optimal dapat diperiksa menggunakan *Davies-Bouldin Index* (DBI). Jumlah kluster yang dipilih adalah jumlah kluster yang memiliki nilai DBI terkecil [15].

2. TINJAUAN PUSTAKA

E-Commerce

E-Commerce adalah proses membeli, menjual, atau memperdagangkan data, barang, atau jasa melalui internet [21]. Media *e-commerce* melibatkan penggunaan *website*, *internet*, dan aplikasi maupun *browser* pada perangkat seluler untuk bertransaksi bisnis. Pada awal berkembangnya *e-commerce*, media digital yang tersedia hanya *web browser*, namun sekarang yang lebih banyak digunakan oleh orang-orang adalah aplikasi pada perangkat seluler [12].

Digital Economy

Pemanfaatan teknologi informasi dan komunikasi terjadi pada hampir semua lini kehidupan termasuk bidang ekonomi. Sektor ekonomi merupakan pionir dalam pemanfaatan internet setelah proyek ARPANET (*Advance Research Projects Agency*) dalam bidang militer. Ekonomi digital pertama kali dikemukakan oleh Tapscott. Tapscott mengungkapkan bahwa internet dan web memunculkan bentuk ekonomi baru berdasarkan jaringan *human intelligence* [20]. Menurut OECD (*Organization for Economic Co-operation and Development*), ekonomi digital adalah pasar yang dibentuk oleh teknologi digital yang memfasilitasi perdagangan barang maupun jasa melalui *e-commerce*.

Analisis Kluster

Analisis kluster merupakan pengelompokan objek menjadi kelompok-kelompok yang lebih kecil, dimana setiap

kelompok berisi objek yang memiliki kemiripan satu sama lain [19]. Analisis kluster adalah kumpulan dari beberapa teknik pengolahan data multivariat yang memiliki tujuan utama mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya [7]. Kluster yang terbentuk harus menunjukkan homogenitas internal yang tinggi dalam satu kluster dan heterogenitas yang tinggi antar kluster.

Farley dan Raftery membagi metode *clustering* ke dalam dua kelompok utama, yaitu metode hirarki dan metode partisi [5]. Algoritma *clustering* hirarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan kluster yang diinginkan belum diketahui banyaknya [13]. Metode hirarki dibagi menjadi dua yaitu, algoritma *agglomerative* (seperti *single-linkage*, *complete-linkage*, *average linkage*, metode *centroid* dan metode *ward*) dan algoritma *divisive*.

Sementara itu, Algoritma *clustering* non hirarki/partisi digunakan untuk pengelompokan objek dimana banyaknya kluster yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengelompokan [11]. Metode partisi merelokasi secara instan data-data yang ada dengan memindahkan data tersebut dari satu *cluster* ke *cluster* yang lain, dan memulai partisi dari awal. Beberapa tipe metode partisi adalah: (i) Error Minimization Algorithms (*K-Means Algorithm*, *K-Modes Algorithm*) dan (ii) *Graph-Theoretic Clustering (Minimal Spanning Tree-MST)* [17].

Uji Multikolinieritas

Pada analisis kluster harus memenuhi asumsi non-multikolinieritas [7]. Multikolinieritas mengindikasikan adanya hubungan linier sempurna atau pasti antara beberapa atau semua variabel [6]. Terkadang variabel independen saling terhubung erat dan sulit untuk membedakan pengaruh masing-masing variabel independen dengan variabel dependen. Keadaan tersebut dinamakan *collinear* atau pada variabel tersebut terjadi masalah kolinieritas. Satu set variabel dikatakan

multikolinier jika terdapat satu atau lebih hubungan linier antara beberapa variabel.

Uji Multikolinieritas digunakan untuk menguji apakah terdapat korelasi antar variabel-variabel bebas. Salah satu cara untuk mendeteksi adanya multikolinieritas adalah dengan menghitung nilai *Varians Inflation Factor (VIF)* [6]. Misalkan terdapat regresi linier :

$$Y = X\beta + \varepsilon$$

atau

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon_k \quad (2.1)$$

X merupakan matriks *full-rank* $n \times p$ dan $\mathcal{L}(\varepsilon) = N(0, \sigma^2 I_n)$. VIF_j dievaluasi dalam langkah-langkah berikut [3]:

1. Jalankan regresi tambahan OLS dari X_j sebagai fungsi dari semua variabel penjelasan lainnya:

$$X_j = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} \dots + \alpha_k X_k + e_i \quad (2.2)$$

$$j = 1, 2, \dots, k$$

2. Mencari nilai VIF untuk setiap j dengan rumus

$$VIF_j = \frac{1}{1-R_j^2} \quad (2.3)$$

dimana R_j^2 adalah koefisien determinasi regresi pada langkah (1)

3. Jika nilai VIF di bawah 10, maka tidak terjadi gejala multikolinieritas [6].

VIF mengukur akibat dari penambahan 1 variabel eksplanatori yang tidak ortogonal pada model regresi linier, dan dapat dihitung sebagai rasio dari determinan [14].

Namun, VIF tidak sepenuhnya dapat digunakan pada model, yang mana mencakup serangkaian regresor, yaitu indikator/dummy untuk variabel kategori yang sama ataupun regresi polinomial. Hal ini terjadi karena korelasi antara variabel-variabel tersebut diinduksi oleh struktur model. Masalahnya adalah bahwa nilai VIF dipengaruhi oleh garis dasar variabel kategoris. Untuk memastikan tidak ada variabel yang memiliki nilai VIF di atas nilai batas yang dapat diterima, perlu untuk mengulangi analisis ini untuk setiap tingkat variabel kategorik. Misal pada suatu variabel umur terdapat 4 kategori yaitu kelompok umur anak-anak, remaja, dewasa, dan lansia, serta kelompok umur anak-

anak sebagai kelompok acuan, maka terdapat tiga dummy variabel untuk kategori umur sehingga kita perlu menghitung VIF dari dummy variabel tersebut (kelompok umur remaja, dewasa, dan lansia). Sehingga untuk mengetahui apakah variabel umur tidak terjadi multikolinearitas perlu dilakukan penghitungan sebanyak banyaknya dummy variabel. Hal ini menjadi lebih rumit jika kita memiliki lebih dari satu variabel kategorik. Maka untuk mengatasi hal tersebut, dapat menggunakan *Generalized VIF* atau GVIF [4]. Dalam menghitung GVIF kita mulai dengan memisahkan variabel dummy untuk variabel kategori umur dalam matriks terpisah X_i dan variabel penjelasan yang tersisa (tidak termasuk konstanta) di $X_{[-i]}$.

GVIF adalah faktor inflasi variasi umum, yang digunakan sebagai pengganti VIF standar karena VIF standar tidak dapat diterapkan pada variabel dengan $Df > 1$ (misalnya, variabel kategorik dengan beberapa kategori). GVIF dihitung untuk set regresor terkait, seperti untuk serangkaian regresor indikator/dummy untuk beberapa jenis variabel kategorik, atau untuk variabel polynomial. Singkatnya, GVIF adalah VIF yang dikoreksi dengan jumlah derajat kebebasan variabel independen, artinya untuk variabel kontinu, VIF dan GVIF adalah sama nilainya. Sedangkan untuk variabel kategorik, kita akan mendapatkan satu nilai GVIF untuk setiap jenis kategori terpisah (misalnya pada variabel umur yang terdiri atas 4 kategori hanya akan menghasilkan satu nilai GVIF). Jadi, variabel yang membutuhkan lebih dari 1 koefisien dan dengan demikian memiliki derajat kebebasan lebih dari 1 biasanya dievaluasi menggunakan GVIF.

Misal model regresi sebagai berikut :

$$Y_{(N \times 1)} = X_{(N \times (l+1))} \beta_{((l+1) \times 1)} + \varepsilon_{(N \times 1)}$$

Atau

$$\begin{matrix} Y \\ (N \times 1) \end{matrix} = \begin{matrix} \beta_0 \\ (N \times r) \end{matrix} + \begin{matrix} X_j \\ (r \times 1) \end{matrix} \begin{matrix} \beta_j \\ (r \times 1) \end{matrix} + \begin{matrix} X_{[-j]} \\ (N \times (l-r)) \end{matrix} \begin{matrix} \beta_{[-j]} \\ ((l-r) \times 1) \end{matrix} + \begin{matrix} \varepsilon \\ (N \times 1) \end{matrix} \quad (2.4)$$

Keterangan :

X_j = Variabel independen yang akan dicari GVIFnya dan suku polinomialnya

$X_{[-j]}$ = Seluruh variabel pada matriks X selain variabel pada X_j , tidak termasuk konstanta

N = Jumlah sampel

l = Jumlah seluruh variabel independen

r = Jumlah variabel indikator/dummy pada variabel kategorik tertentu

Maka GVIF didefinisikan sebagai berikut:

$$GVIF_j = \frac{\det(R_{X_j}) \det(R_{X_{[-j]}})}{\det(R)} \quad (2.5)$$

Dimana :

R_{X_j} = Matriks korelasi untuk X_j

$R_{X_{[-j]}}$ = Matriks korelasi untuk $X_{[-j]}$

R = Matriks korelasi untuk seluruh variabel pada matriks X, kecuali konstanta

Untuk membuat GVIF sebanding di seluruh dimensi dan mengurangi GVIF ke ukuran linier Fox dan Monette juga menyarankan menggunakan $GVIF^{\frac{1}{2 \times df}}$ [4]. Dimana df adalah jumlah koefisien dalam subset variabel (misalnya, untuk variabel umur dengan kategori umur anak-anak sebagai variabel acuan, $df = 3$). Kemudian kita dapat mendeteksi multikolinearitas dengan aturan yang sama dengan VIF yaitu nilai kurang dari 10 menunjukkan tidak adanya multikolinearitas.

Algoritma K-Modes Clustering

Proses klusterisasi pada data kategorik (nominal atau ordinal) yang tidak dapat menggunakan algoritma *k-means clustering* membuat tercetusnya suatu algoritma baru pada tahun 1997 yang disebut dengan *k-modes clustering*. Algoritma data mining ini merupakan pengembangan dari algoritma *k-means clustering* dan digunakan untuk mengelompokkan data kategorik dengan menghitung fungsi jarak *mismatching* [23]. Hal ini yang membedakan penggunaan algoritma *k-means* dengan *k-modes*. Penggunaan *k-means clustering* hanya terbatas pada jenis data numerik (interval/rasio) yang pengklusterisasian datanya dilakukan dengan menghitung fungsi jarak *euclidean* dan rata-rata suatu data dengan data yang lain serta tidak dapat mengklusterisasikan data kategorik. Sementara itu, *k-modes clustering* dapat melakukan klusterisasi pada data kategorik dengan cara menghitung ukuran ketidakmiripan antara dua data [23].

Tahapan *clustering* menggunakan algoritma *k-modes* adalah sebagai berikut [2] :

1. Menghitung jumlah kluster (k) yang akan dibentuk dari r data. Banyaknya kluster harus lebih sedikit dari banyaknya data ($k < r$).
2. Memilih k modus awal dari data secara acak sebagai pusat kluster (*centroid*). Jumlah *centroid* yang dibuat harus sama dengan jumlah kluster yang akan dibentuk. Misalnya, jika ingin membentuk tiga kluster, maka *centroid* juga berjumlah tiga.
3. Menghitung jarak setiap data (objek) terhadap semua *centroid*.

Jarak dihitung dengan menggunakan ukuran ketidakmiripan sederhana (*simple matching dissimilarity*) Asumsikan A_1 dan A_2 adalah dua data yang memiliki n variabel kategorik. Ukuran ketidakmiripan antara A_1 dan A_2 dapat didefinisikan sebagai total ketidakcocokan dari variabel-variabel kategorik yang sesuai dari kedua data. Semakin kecil nilai ketidakcocokan, semakin mirip kedua data tersebut [10]. Persamaan ukuran ketidakmiripan sederhana dirumuskan sebagai berikut.

$$d(A_1, A_2) = \sum_{j=1}^n \delta(x_{1j}, x_{2j}) \quad (2.6)$$

dengan

$$\delta(x_{1j}, x_{2j}) = \begin{cases} 0 & x_{1j} = x_{2j} \\ 1 & x_{1j} \neq x_{2j} \end{cases} \quad (2.7)$$

x_{1j} adalah nilai dari variabel ke- j pada objek A_1 dan x_{2j} adalah nilai dari variabel ke- j pada objek A_2 .

4. Mengalokasikan objek berdasarkan jarak terdekat ke *centroid* dengan rumus jarak terdekat. Nilai ukuran ketidakmiripan sederhana ini dapat menunjukkan kluster mana yang paling dekat dengan data.
5. Jika semua data telah dialokasikan ke dalam kluster terdekat, selanjutnya memperbaiki *centroid* masing-masing kluster berdasarkan modus dari setiap variabel anggota kluster yang terbentuk.
6. Menghitung ulang jarak setiap data terhadap semua *centroid* baru dengan menggunakan ukuran ketidakmiripan sederhana seperti langkah (3). Jika pada penghitungan ulang jarak, data lebih

mendekati kluster lain daripada kluster saat ini, maka data dialokasikan ulang ke kluster lain yang lebih dekat tersebut.

7. Mengulangi langkah (5) dan (6) hingga tidak ada data yang berpindah kluster.

Modus Kluster

Pada algoritma *k-modes clustering*, pusat kluster (*centroid*) diwakili oleh vektor modus dari variabel kategorik. Dalam statistika, modus dari sekumpulan nilai yaitu nilai yang paling sering muncul. Jika sekumpulan data memiliki n variabel kategorik. Vektor modus V terdiri dari n nilai kategorik (v_1, v_2, \dots, v_n) yang masing-masing merupakan modus dari sebuah variabel. Vektor modus dari sebuah kluster meminimalkan jumlah jarak antar setiap data di dalam karakter dengan pusat kluster [9].

Indeks Davies-Bouldin

Davies-Bouldin Index (DBI) adalah suatu ukuran yang digunakan untuk menentukan jumlah kluster yang terbaik setelah proses pengklusteran selesai [15]. Pendekatan DBI ini bertujuan untuk memaksimalkan jarak antara kluster yang satu dengan kluster yang lain dan mencoba meminimalkan jarak antar objek dalam suatu kluster [8]. Semakin kecil nilai DBI yang diperoleh ($\text{non-negatif} \geq 0$) maka semakin baik kluster yang diperoleh dari pengelompokan *k-modes clustering* yang digunakan [18]. Formula Indeks Davies-Bouldin adalah sebagai berikut.

$$DBI = \frac{1}{k} \times \sum_{a=1}^k R_a \quad (2.8)$$

dengan

$$R_a = \max_{a \neq b} R_{ab} \text{ dan } R_{ab} = \frac{s_a + s_b}{d(v_a, v_b)} \quad (2.9)$$

Keterangan :

DBI = Indeks Davies-Bouldin

k = Jumlah kluster

R_{ab} = Ukuran kemiripan antara kluster ke- a dan kluster ke- b

s_a = Ukuran dispersi kluster ke- a

s_b = Ukuran dispersi kluster ke- b

a = 1, 2, ..., k

b = 1, 2, ..., k

$$s_a = \left[\frac{1}{n_a} \sum_{T_i \in c_a, i=1}^{n_a} (d(T_i, V_a))^2 \right]^{\frac{1}{2}} \quad (2.10)$$

Keterangan :

n_a = Banyaknya anggota kluster ke-a

c_a = Kluster ke-a

T_i = Anggota ke-i pada kluster ke-a

V_a = *Centroid* kluster ke-a

$d(T_i, V_a)$ = Jarak dari anggota ke-i pada kluster ke-a (T_i) dengan *centroid* kluster ke-a (V_a)

Nilai $d(T_i, V_a)$ dihitung menggunakan ukuran ketidaksamaan percocokan sederhana yang dapat dijabarkan sebagai berikut :

$$d(T_i, V_a) = \sum_{j=1}^n \delta(x_{ij}, v_{aj}) \quad (2.11)$$

dengan

$$\delta(x_{ij}, v_{aj}) = \begin{cases} 0 & x_{ij} = v_{aj} \\ 1 & x_{ij} \neq v_{aj} \end{cases} \quad (2.12)$$

Dimana

x_{ij} = Nilai dari variabel ke-j pada T ke-i

v_{aj} = Nilai ke-j pada *centroid* kluster ke-a

N = Jumlah variabel

Formula untuk mencari nilai varians kluster ke-b juga sama dengan formula untuk mencari nilai varians kluster ke-a di atas.

Ukuran ketidakmiripan percocokan sederhana dapat juga digunakan untuk menghitung jarak *centroid* kluster ke-a (V_a) ke *centroid* kluster ke-b (V_b) sebagai berikut :

$$d(V_a, V_b) = \sum_{j=1}^n \delta(v_{aj}, v_{bj}) \quad (2.13)$$

dengan

$$\delta(v_{aj}, v_{bj}) = \begin{cases} 0 & v_{aj} = v_{bj} \\ 1 & v_{aj} \neq v_{bj} \end{cases} \quad (2.14)$$

dimana v_{bj} adalah nilai ke-j pada *centroid* kluster ke-b.

3. METODOLOGI

Penelitian ini menggunakan data sekunder yang diperoleh dari *raw data* hasil survei Praktik Kerja Lapangan (PKL) Politeknik Statistika STIS Tahun Akademik 2019/2020. Salah satu tujuan dari survei tersebut adalah

untuk mengetahui karakteristik pelaku ekonomi digital di Daerah Istimewa Yogyakarta. Pengambilan data pada survei menerapkan *probability sampling* dengan desain *Stratified Two Stage Sampling*. Variabel yang digunakan pada penelitian ini adalah jenis kelamin, umur dan jenis platform yang digunakan dalam melakukan pembelian barang secara *online*. Jenis platform yang dicakup yaitu Marketplace (Shopee, Bukalapak, Tokopedia, Lazada, lainnya) dan Media Sosial (Instagram, Facebook, Media Chat, Twitter).

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel
X ₁	Jenis Kelamin
X ₂	Umur
X ₃	Shopee
X ₄	Bukalapak
X ₅	Tokopedia
X ₆	Lazada
X ₇	Marketplace Lainnya
X ₈	Instagram
X ₉	Facebook
X ₁₀	Media Chat
X ₁₁	Twitter

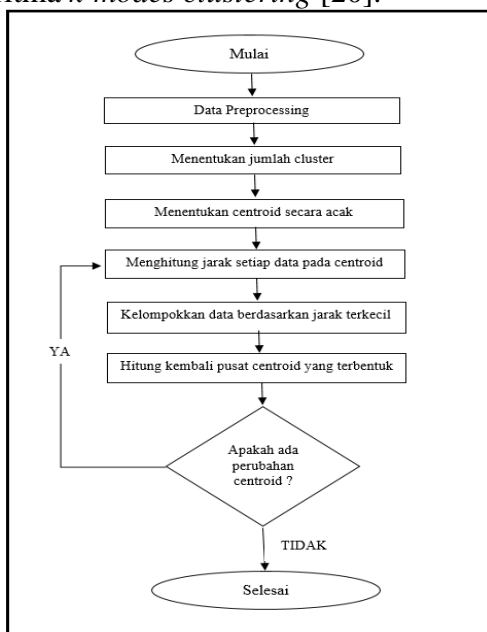
Pada variabel jenis kelamin terdapat dua kategori yaitu Laki-laki = 1 dan Perempuan = 2; variabel umur terdiri atas tiga kategori yaitu Muda = 1, Dewasa = 2, dan Tua = 3; variabel shopee, bukalapak, sampai twitter memiliki dua kategori yaitu Ya = 1 dan Tidak = 0.

Metode analisis yang dilakukan berupa analisis secara deskriptif maupun inferensia. Analisis deskriptif menggambarkan karakteristik pelaku ekonomi digital untuk masing-masing variabel menggunakan diagram lingkaran. Selanjutnya, secara inferensia menggunakan metode *k-modes clustering* untuk mengelompokkan minat belanja *online* berdasarkan variabel yang digunakan. Data yang diambil sebanyak 2415 *record* dilakukan *preprocessing* terlebih dahulu untuk mengatasi permasalahan yang dapat mengganggu proses pengolahan data. Setelah itu, dilakukan uji asumsi non-multikolinearitas untuk membuktikan bahwa tidak terdapat kolinearitas antar variabel.

Kemudian, langkah-langkah yang dilakukan untuk analisis kluster pada penelitian ini adalah sebagai berikut :

1. Memasukkan data variabel
2. Menentukan nilai $k = 2,3,\dots,10$
3. Menentukan *centroid* (pusat kluster) secara acak dari masing-masing kluster
4. Menghitung jarak tiap objek terhadap *centroid* menggunakan ukuran ketidaksamaan sederhana.
5. Mengelompokkan objek berdasarkan jarak terdekat ke *centroid*.
6. Menentukan modus dari setiap variabel sebagai *centroid* kluster.
7. Menghitung ulang jarak tiap objek terhadap *centroid* baru menggunakan ukuran ketidaksamaan sederhana.
8. Mengelompokkan objek berdasarkan jarak terdekat ke *centroid*.
9. Apabila ada objek yang berpindah kluster, maka ulangi langkah ke enam sampai delapan hingga tidak ada objek yang berpindah kluster.
10. Memilih k terbaik menggunakan *Davies Bouldin Index*.
11. Membuat profiling dari masing-masing kluster.

Gambar 3.1 merupakan diagram alir algoritma *k-modes clustering* [20].



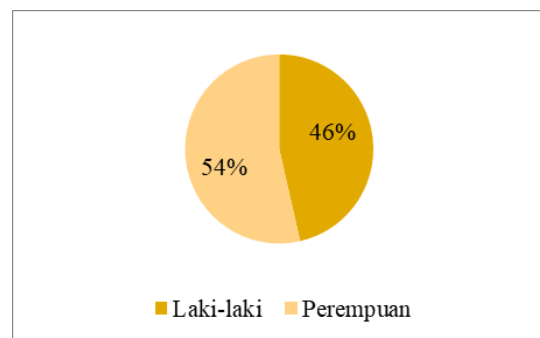
Gambar 3.1. Flowchart Proses Klustering

4. PEMBAHASAN

Analisis Deskriptif

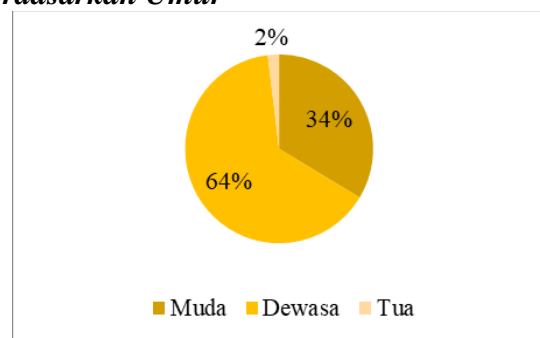
Persentase Penduduk Daerah Istimewa Yogyakarta yang Melakukan Belanja Online Berdasarkan Jenis Kelamin

Berdasarkan gambar 4.1 diketahui bahwa persentase penduduk perempuan yang melakukan kegiatan belanja *online* adalah sebesar 54 persen. Sedangkan untuk penduduk laki-laki adalah sebesar 46 persen. Hal tersebut menunjukkan bahwa pada Daerah Istimewa Yogyakarta kecenderungan belanja *online* dilakukan oleh penduduk perempuan maupun laki-laki.



Gambar 4.1. Persentase Penduduk yang melakukan Belanja *Online* Berdasarkan Jenis Kelamin

Persentase Penduduk Daerah Istimewa Yogyakarta yang Melakukan Belanja Online Berdasarkan Umur

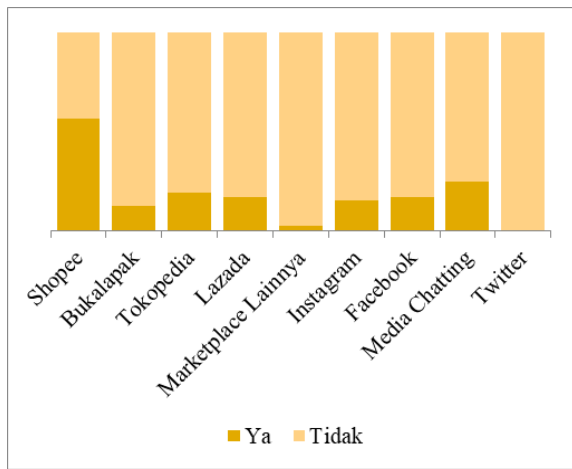


Gambar 4.2 Persentase Penduduk yang melakukan Belanja *Online* Berdasarkan Umur

Berdasarkan gambar 4.2 penduduk pada Daerah Istimewa Yogyakarta 64 persen yang melakukan belanja *online* merupakan umur

dewasa yaitu 25-59 tahun. Untuk umur muda yang melakukan kegiatan belanja *online* adalah sebesar 34 persen dan untuk 2 persen sisanya adalah umur tua. Hal ini menunjukkan bahwa mayoritas penduduk yang melakukan kegiatan belanja *online* adalah umur dewasa yang notabenehnya sudah memiliki pekerjaan dan juga penghasilan. Umur tersebut merupakan umur produktif bekerja.

Persentase Penduduk Daerah Istimewa Yogyakarta yang Melakukan Belanja Online Berdasarkan Platform



Gambar 4.3 Persentase Penduduk yang melakukan Belanja *Online* Berdasarkan Platform

Jika dilihat berdasarkan platform yang digunakan, mayoritas penduduk di Daerah Istimewa Yogyakarta menggunakan aplikasi *Shopee* sebagai media untuk belanja *online*, yaitu sebesar 56,36 persen. Sedangkan platform yang paling sedikit digunakan untuk belanja *online* adalah *Twitter*.

ANALISIS INFERENSIA

Uji Multikolinearitas

Dalam melakukan analisis kluster, terdapat beberapa asumsi yang harus dipenuhi yaitu sampel representatif (mewakili populasi) dan non-multikolinearitas (tidak ada korelasi antar variabel). Pengujian untuk asumsi non-multikolinearitas dapat diketahui dari nilai *Generalized Variance Inflation Factor* (GVIF), jika nilai GVIF < 10 maka asumsi non-multikolinearitas terpenuhi.

Tabel 4.1 Nilai GVIF Masing-Masing Variabel

Variabel	$\frac{1}{GVIF_{2 \times df}}$
Jenis Kelamin	1,0584
Umur	1,0102
Shopee	1,0985
Bukalapak	1,0636
Tokopedia	1,0542
Lazada	1,0346
Marketplace Lainnya	1,0107
Instagram	1,0964
Facebook	1,1046
Media Chatting	1,1534
Twitter	1,0092

Berdasarkan Tabel 2 dapat diketahui bahwa nilai GVIF masing-masing variabel bernilai kurang dari 10 sehingga dapat disimpulkan bahwa asumsi non-multikolinearitas terpenuhi.

Hasil Pengklasteran

Setelah dilakukan klusterisasi terhadap daa, didapatkan hasil yang ditunjukkan pada tabel 4.2 untuk *centroid* atau modus dari masing-masing kluster dan masing-masing k.

Tabel 4.2 Hasil pengklasteran k = 2,3,...,10

k	Kluster	Jumlah Anggota (n _a)	Centroid (V _a)
2	1	1595	1,2,0,0,0,0,0,0,0,0
	2	820	2,2,1,0,0,0,0,0,0,0
3	1	1361	1,2,0,0,0,0,0,0,0,0
	2	787	2,2,1,0,0,0,0,0,0,0
	3	267	1,1,0,0,0,0,0,1,0,1
4	1	667	2,2,0,0,0,0,0,0,0,1
	2	1146	2,2,1,0,0,0,0,0,0,0
	3	203	1,1,0,0,0,0,0,1,0,1
	4	399	1,2,0,0,0,0,0,0,0,0
5	1	663	2,2,0,0,0,0,0,0,0,1
	2	1134	2,2,1,0,0,0,0,0,0,0
	3	80	1,1,0,0,0,0,0,1,0,1
	4	399	1,2,0,0,0,0,0,0,0,0
	5	119	1,1,1,1,1,0,0,1,1,1
6	1	624	2,2,0,0,0,0,0,0,0,1
	2	1134	2,2,1,0,0,0,0,0,0,0
	3	80	2,1,0,0,0,0,0,1,0,1
	4	399	1,2,0,0,0,0,0,0,0,0
	5	104	1,1,1,1,1,0,0,1,1,1
	6	74	1,2,0,0,0,0,0,0,1,1

	1	612	2,2,0,0,0,0,0,0,1,0
	2	1006	2,2,1,0,0,0,0,0,0,0
	3	80	2,1,0,0,0,0,0,1,0,1,0
7	4	380	1,2,0,0,0,0,0,0,0,0,0
	5	85	1,1,1,1,1,0,0,0,1,1,1,0
	6	74	1,2,0,0,0,0,0,0,0,1,1,0
	7	178	1,2,1,1,1,0,0,0,0,0,0,0
	1	545	2,2,0,0,0,0,0,0,0,1,0
	2	996	2,2,1,0,0,0,0,0,0,0,0
	3	69	2,1,0,0,0,0,0,0,1,0,1,0
8	4	380	1,2,0,0,0,0,0,0,0,0,0
	5	77	1,1,1,1,1,0,0,0,1,1,1,0
	6	71	1,2,0,0,0,0,0,0,0,1,1,0
	7	178	1,2,1,1,1,0,0,0,0,0,0,0
	8	99	2,2,1,0,0,0,0,0,1,1,1,0
	1	516	2,2,0,0,0,0,0,0,0,0,1,0
	2	979	2,2,1,0,0,0,0,0,0,0,0,0
	3	67	2,1,0,0,0,0,0,0,1,0,1,0
	4	335	1,2,0,0,0,0,0,0,0,0,0,0
9	5	76	1,1,1,1,1,0,0,0,1,1,1,0
	6	71	1,2,0,0,0,0,0,0,0,1,1,0
	7	175	1,2,1,1,1,0,0,0,0,0,0,0
	8	99	2,2,1,0,0,0,0,0,1,1,1,0
	9	97	1,1,0,0,0,1,0,0,0,0,0,0
	1	359	1,2,0,0,0,0,0,0,0,0,1,0
	2	411	1,2,1,0,0,0,0,0,0,0,0,0
	3	314	2,2,0,0,0,0,0,0,1,0,1,0
	4	256	1,2,0,0,0,0,0,0,0,0,0,0
10	5	75	1,1,1,1,1,0,0,0,1,1,1,0
	6	150	2,2,0,0,0,0,0,0,0,1,1,0
	7	103	1,2,1,1,1,0,0,0,0,0,0,0
	8	122	2,2,1,0,0,0,0,0,1,1,1,0
	9	67	1,1,0,0,0,1,0,0,0,0,0,0
	10	558	2,2,1,0,0,0,0,0,0,0,0,0

(i). Jarak masing-masing anggota kluster 1 terhadap *centroid* kluster 1

$$d(T_1, V_1) = \delta(1,1) + \delta(1,2) + \delta(0,0) + \delta(0,0) + \delta(1,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) = 0 + 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 = 2$$

∴

Dan seterusnya sampai anggota ke-1595

(ii). Jarak masing-masing anggota kluster 2 terhadap *centroid* kluster 2

$$d(T_2, V_2) = \delta(2,2) + \delta(2,2) + \delta(1,1) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(1,0) + \delta(1,0) + \delta(0,0) + \delta(0,0) = 0 + 1 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 = 2$$

∴

Dan seterusnya sampai anggota ke-820

3. Ukuran dispersi masing-masing kluster

Ukuran dispersi kluster 1

$$s_1 = \left[\frac{1}{1595} \times (2^2 + 1^2 + 5^2 + \dots + 4^2) \right]^{\frac{1}{2}} = 2,4817$$

Ukuran dispersi kluster 2

$$s_2 = \left[\frac{1}{820} \times (2^2 + 1^2 + 3^2 + \dots + 3^2) \right]^{\frac{1}{2}} = 1,6770$$

4. Ukuran kemiripan antara kluster 1 dan 2

$$R_{12} = \frac{s_1 + s_2}{d(V_1, V_2)} = \frac{2,4817 + 1,6770}{2} = 2,0794$$

5. Matriks R (2x2)

$$R = \begin{bmatrix} - & 2,0794 \\ 2,0794 & - \end{bmatrix}$$

6. Nilai *Davies-Bouldin Index*

$$DBI = \frac{1}{2} \sum_{a=1}^2 R_a = \frac{1}{2} \times (2,0794 + 2,0794) = 2,0794$$

Langkah yang sama dilakukan untuk mencari DBI k=3,4,...,10 sehingga diperoleh nilai DBI yang tertera pada tabel 3.

Tabel 4.3 Nilai DBI k=2,3,...,10

k	DBI
2	2,0794
3	1,7293
4	1,8206
5	1,6645
6	1,6264
7	1,5236
8	1,4067

Penentuan Jumlah Kluster Optimal

Jumlah kluster yang optimal ditentukan menggunakan metode *Davies-Bouldin Index* (DBI). Semakin kecil nilai DBI akan memberikan hasil yang baik, berikut merupakan contoh penghitungan nilai DBI untuk k=2.

1. Jarak antar *centroid* kluster

Jarak antara *centroid* kluster 1 dan 2

$$d(V_1, V_2) = \delta(1,2) + \delta(2,2) + \delta(0,1) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) + \delta(0,0) = 1 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 2$$

2. Jarak masing-masing anggota kluster terhadap *centroid* masing-masing kluster

9	1,3427
10	1,6398

Pada Tabel 4.3 dapat dilihat bahwa DBI terkecil dimiliki oleh $k=9$ yaitu sebesar 1,3427.

Sehingga dapat disimpulkan bahwa jumlah kluster optimal untuk mengelompokkan minat belanja *online* sebanyak 9 kluster.

Interpretasi dan Profiling Hasil Kluster $k=9$

Tabel 4.4 Karakteristik masing-masing kluster untuk $k=9$

Kluster	JK	Umur	Shopee	Bukalapak	Tokopedia	Lazada	Lainnya	Ig	Fb	Media Chatting	Twitter	Jumlah Anggota
1	Perempuan	Dewasa	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	516
2	Perempuan	Dewasa	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	979
3	Perempuan	Muda	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	67
4	Laki-laki	Dewasa	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	335
5	Laki-laki	Muda	Ya	Ya	Ya	Tidak	Tidak	Ya	Ya	Ya	Tidak	76
6	Laki-laki	Dewasa	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Tidak	71
7	Laki-laki	Dewasa	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	175
8	Perempuan	Dewasa	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Ya	Tidak	99
9	Laki-laki	Muda	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	97

Berdasarkan Tabel 4.4, berikut adalah penamaan dan karakteristik untuk masing-masing kluster

- a. Kluster 1, cukup minat belanja melalui media sosial

Kluster 1 terdiri atas orang-orang yang cukup minat belanja melalui Media Sosial yaitu Media Chatting, kategori ini didominasi oleh perempuan umur dewasa (25-59 tahun)

- b. Kluster 2, cukup minat belanja melalui Marketplace

Kluster 2 terdiri atas orang-orang yang cukup minat belanja melalui Marketplace yaitu Shopee, kategori ini didominasi oleh perempuan umur dewasa (25-59 tahun)

- c. Kluster 3, minat belanja melalui Media Sosial

Kluster 3 terdiri atas orang-orang yang minat belanja melalui Media Sosial yaitu Instagram dan Media Chatting, kategori ini didominasi oleh perempuan umur muda (15-24 tahun)

- d. Kluster 4, tidak minat belanja *online*

Kluster 4 terdiri atas orang-orang yang tidak minat belanja melalui Marketplace maupun Media Sosial, kategori ini didominasi oleh laki-laki umur dewasa (25-59 tahun)

- e. Kluster 5, sangat minat belanja melalui Marketplace dan Media Sosial

Kluster 5 terdiri atas orang-orang yang sangat minat belanja melalui Marketplace yaitu Shopee, Bukalapak, dan Tokopedia, serta Media Sosial yaitu Instagram, Facebook, dan Media Chatting, kategori ini didominasi oleh laki-laki umur muda (15-24 tahun)

- f. Kluster 6, minat belanja melalui Media Sosial

Kluster 6 terdiri atas orang-orang yang minat belanja melalui Media Sosial yaitu Facebook dan Media Chatting, kategori ini didominasi oleh laki-laki umur dewasa (25-59 tahun)

- g. Kluster 7, sangat minat belanja melalui Marketplace

Klaster 7 terdiri atas orang-orang yang sangat minat belanja melalui Marketplace yaitu Shopee, Bukalapak, dan Tokopedia, kategori ini didominasi oleh laki-laki umur dewasa (25-59 tahun)

- h. Klaster 8, cukup minat belanja melalui Marketplace dan sangat minat belanja melalui media sosial

Klaster 8 terdiri atas orang-orang yang cukup minat belanja melalui Marketplace yaitu Shopee, serta sangat minat belanja melalui Media Sosial yaitu Instagram, Facebook, dan Media Chatting, kategori ini didominasi oleh perempuan umur dewasa (25-59 tahun)

- i. Klaster 9, cukup minat belanja melalui Marketplace

Klaster 9 terdiri atas orang-orang yang cukup minat belanja melalui Marketplace yaitu Lazada, kategori ini didominasi oleh laki-laki umur muda (15-24 tahun)

Sebagian besar masyarakat di DI Yogyakarta termasuk dalam kategori cukup minat belanja melalui Marketplace yaitu Shopee. Masyarakat yang tidak minat melakukan belanja *online* juga masih cukup banyak. Apabila kesembilan klaster dibandingkan, maka klaster yang terbaik adalah klaster 5 dimana merupakan klaster yang anggotanya sangat minat belanja melalui Marketplace dan Media Sosial. Marketplace yang diminati adalah Shopee, Bukalapak, dan Tokopedia, sedangkan Media Sosial yang diminati adalah Instagram, Facebook, dan Media Chatting. Kemudian laki-laki umur muda (15-24 tahun) yang mendominasi kategori ini. Sehingga bagi pengusaha yang ingin membuka *online shop* atau bahkan sudah memiliki *online shop* disarankan untuk menjual barang-barang laki-laki untuk umur muda dan memasarkannya melalui Shopee, Bukalapak, Tokopedia, Instagram, Facebook, atau Media Chatting sehingga lebih banyak memperoleh pembeli.

5. KESIMPULAN

Berdasarkan analisis deskriptif dapat disimpulkan bahwa penduduk Daerah Istimewa Yogyakarta baik laki-laki maupun perempuan memiliki kecenderungan yang sama dalam melakukan belanja secara *online*, dimana penduduk yang paling banyak melakukan belanja *online* berumur dewasa (25-59 tahun). Hal ini karena penduduk yang masuk dalam rentang umur ini biasanya sudah bekerja dan memiliki penghasilan. Platform yang sering digunakan untuk belanja *online* adalah Shopee sedangkan paling sedikit digunakan adalah Twitter.

Berdasarkan hasil DBI, jumlah klaster optimum untuk mengelompokkan minat belanja *online* penduduk Daerah Istimewa Yogyakarta adalah 9 klaster dimana klaster 5 merupakan klaster dengan kategori terbaik dengan yaitu sangat minat belanja melalui Marketplace dan Media Sosial. Marketplace yang diminati adalah Shopee, Bukalapak, dan Tokopedia, sedangkan Media Sosial yang diminati adalah Instagram, Facebook, dan Media Chatting. Klaster ini didominasi oleh laki-laki umur muda (15-24 tahun).

Berdasarkan hal tersebut disarankan bagi pengusaha yang ingin membuka *online shop* atau bahkan sudah memiliki *online shop* disarankan untuk menjual barang-barang laki-laki untuk umur muda dan memasarkannya melalui Shopee, Bukalapak, Tokopedia, Instagram, Facebook, atau Media Chatting sehingga lebih banyak memperoleh pembeli.

6. DAFTAR PUSTAKA

- [1] Anita, Muharmi, Y., & Siska, S. T. (2018). Penentuan Tingkat Minat Belanja Online Melalui Media Sosial Menggunakan Metode Clustering K-Means. *Rang Teknik Journal Vol. 1 No. 2*, 126-134.
- [2] Badruttamam, A., Sudarno, & Maruddani, D. A. I. (2020). Penerapan Analisis Klaster k-Modes dengan Validasi Davies Bouldin Index dalam Menentukan Karakteristik

- Kanal Youtube di Indonesia. *Jurnal Gaussian Vol.9 No.3*, 263-272.
- [3] Buteikis, A. (2018) Practical Econometric and Data Science. Diakses pada 11 Januari 2021, dari http://web.vu.lt/mif/a.buteikis/wp-content/uploads/PE_Book/1-intro.html.
- [4] Fox, J. & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association Vol. 87 No. 417*, 178–83.
- [5] Fraley, C. & Raftery A. E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal Vol.41 No.8*, 578-588.
- [6] Gujarati, D. N. & Porter, D. C. (2009). *Basic Econometrics Fifth Edition*. New York : McGraw-Hill.
- [7] Hair, J. F., et al. (2014). *Multivariate Data Analysis Seventh Edition*. Harlow : Pearson.
- [8] Hilmi, M. N., Wilandari, Y., & Yasin, H. (2015). Pemetaan Preferensi Mahasiswa Baru dalam Memilih Jurusan Menggunakan Artificial Neural Network (ANN) dengan Algoritma Self Organizing Maps (SOM). *Jurnal Gaussian Vol. 4 N. 1*, 53-60.
- [9] Huang, J. Z. (2009). Clustering Categorical Data with k-Modes. *Encyclopedia of Data Warehousing and Mining*.
- [10] Huang, Z. (1997). A Fast Clustering Algorithm to Cluster very Large Categorical Data Sets in Data Mining. *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'97)*.
- [11] Johnson, R. A. & Winchern, D. W. (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- [12] Laudon, K. & Traver, C. G. (2017). *E-Commerce 2016 : Business, Technology, Society*. Harlow : Pearson Education.
- [13] Mattjik, A.A. & Sumertajaya, I.M. (2011). *Sidik Peubah Ganda*. Bogor : IPB Press.
- [14] O'Driscoll, D. & Ramirez, D. E. (2015). Response Surface Designs Using the Generalized Variance Inflation Factors. *Cogent Mathematics Vol. 2 No. 1*.
- [15] Permatadevi, M. A., Hendrawan, R. A., & Hafidz, I. (2013). Karakteristik Pelanggan Telepon Kabel Menggunakan Clustering SOM dan K-Means untuk Mengurangi Kesalahan Klasifikasi Pelanggan Perusahaan Telekomunikasi. *Jurnal Teknik Pomits Vol.1 No.1*, 1-6.
- [16] Ramdhani, F., Hoyyi, A., & Mukid, M. A. (2015). Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster. *Jurnal Gaussian Vol.4 No.4*, 875-884.
- [17] Romzi, M., Anggorowati, M.A., Mariyah, S. Data Mining And Knowledge Management. Jakarta : Sekolah Tinggi Ilmu Statistik.
- [18] Sukamto, Daqiqil, I., & Angraini, T. R. (2018). Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means. *Jurnal Informatika Vol.6 No.2*, 137-148.
- [19] Supranto, J. (2004). *Analisis Multivariat : Arti dan Interpretasi*. Jakarta : Rineka Cipta.
- [20] Tapscott, D. (2015). *The Digital Economy Anniversary Edition : Rethinking Promise and Peril in the Age of Networked Intelligence*. New York : McGraw-Hill Education.
- [21] Turban, E., et al. (2015). *Electronic Commerce : a Managerial and Social Networks Perspective Eighth Edition*. Springer Text in Business and Economics.
- [22] Van Ark, B., et al. (2016). *Navigating the new digital economy: driving digital growth and productivity from installation to deployment*.
- [23] Yuliana, L. (2019). *Penerapan Algoritma K-Modes Clustering untuk Pengelompokan Desa Rawan Kebakaran di Provinsi Riau (Studi Kasus : BPBD Riau)*. Tugas Akhir. Pekanbaru : Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.
- [24] Yulianita, T., & Istiawan, D. (2017). Implementasi Algoritma K-Modes untuk

Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis. The 6th University Research Colloquium 2017 Universitas Muhammadiyah Magelang, 429-440