

## CLUSTERING TINGKAT RISIKO KLASIFIKASI LAPANGAN USAHA (KLU) MENGUNAKAN METODE K-MEANS

Jessica Rahmawati Nugroho\*<sup>1</sup>, Yoyon K. Suprpto<sup>2</sup>, Eko Setijadi<sup>3</sup>

<sup>1,2,3</sup>Institut Teknologi Sepuluh Nopember, Surabaya  
Email: <sup>1</sup>nugrohojessica@gmail.com, <sup>2</sup>yoyonsuprpto@ee.its.ac.id, <sup>3</sup>ekoset@ee.its.ac.id  
\*Penulis Korespondensi

(Naskah masuk: 29 Juni 2021, diterima untuk diterbitkan: 31 Mei 2022)

### Abstrak

Pajak adalah sumber utama pendapatan negara. Karena itu, otoritas pajak di seluruh dunia, bertugas untuk mengurangi kesenjangan pajak (*tax gap*). Salah satu faktor yang menyebabkan adanya kesenjangan pajak adalah tingkat kepatuhan Wajib Pajak (WP). Dalam upaya meminimalisir ketidakpatuhan WP, Direktorat Jenderal Pajak (DJP) melakukan kegiatan pengawasan dan pemeriksaan terhadap WP. WP badan mempunyai kontribusi dominan terhadap penerimaan negara. Dari latar belakang tersebut, diambil pendekatan menggunakan metode *clustering* dengan algoritma K-Means untuk mengelompokkan Klasifikasi Lapangan Usaha (KLU) sesuai risiko tingkat ketidakpatuhan dan dampak fiskal bagi penerimaan yang dibagi menjadi tinggi, sedang, dan rendah. Hasil pengujian menunjukkan terdapat 9 KLU yang masuk ke dalam kuadran berwarna merah dengan tingkat ketidakpatuhan tinggi (variabel A) dan memiliki dampak fiskal yang tinggi (variabel B). Dengan validasi *clustering* menggunakan uji *silhouette*, diperoleh nilai 0,65 untuk variabel A dan 0,93 untuk variabel B. Informasi yang dihasilkan dari penelitian ini dapat dipergunakan untuk mendukung pengambilan keputusan dalam penentuan daftar KLU yang perlu diprioritaskan untuk dilakukan pemeriksaan dan pengawasan.

**Kata kunci:** *Clustering, K-Means, Klasifikasi Lapangan Usaha, Pajak, Koefisien silhouette*

## CLUSTERING RISK LEVEL OF BUSINESS CLASSIFICATIONS USING K MEANS

### Abstract

Tax are the main source of state revenue. Therefore, tax authorities around the world are in charged to reducing the tax gap. One of the factors that causes the tax gap is the level of taxpayer compliance. In order to minimising the risk of taxpayers non-compliance, Directorate General of Taxes (DGT) as a Tax Authorities need to supervising and inspecting taxpayers. Corporate taxpayers as one of the largest source of revenue, based on their business sector have a dominant contribution to state revenues. Thus, in this study, the researcher tries to implement the clustering method with the K-Means algorithm to grouping the business classifications (Klasifikasi Lapangan Usaha/KLU) according to the risk of non-compliance level and the fiscal impact on revenue which is divided into high, medium, and low. The results show that there are 9 KLUs that included into the red quadrant with a high level of non-compliance (variable A) and have a high fiscal impact (variable B). Clustering validation using the silhouette test, obtained values of variable A and variable B, respectively 0,65 and 0,93. The information provided from this study can be used to support decision making in determining the list of KLUs that need to be prioritized for supervising and inspecting.

**Keywords:** *Clustering, K-Means, Business Classifications, Tax, Silhouette coefficient*

### 1. PENDAHULUAN

Adanya *tax gap* pada suatu negara membuat penerimaan negara menjadi tidak optimal. *Tax gap* adalah perbedaan antara jumlah pajak yang harus dibayar ke negara, dan apa yang sebenarnya dibayar (Subroto, 2020).

Salah satu faktor yang menyebabkan adanya *tax gap* adalah tingkat kepatuhan Wajib Pajak (WP) (Raczkowski & Mróz, 2018). Kepatuhan WP

memerlukan kesadaran pembayar pajak terhadap norma dan aturan perundang-undangan yang berlaku dalam pemenuhan kewajiban perpajakan. Sedangkan WP yang tidak patuh, cenderung akan melakukan kegiatan penghindaran pajak baik secara formal maupun informal. Termasuk dalam penghindaran pajak, yaitu pengurangan pajak yang legal secara hukum, maupun penghindaran pajak dengan melakukan pelalaian pembayaran pajak. Pada akhirnya serangkaian kegiatan tersebut akan

merugikan negara dan menyebabkan penerimaan pajak negara berkurang (Hanifah & Agung S, 2016).

Mengingat penerapan *Self Assessment* pada sistem perpajakan Indonesia, dimana WP dipercaya untuk menghitung, membayar, dan melaporkan pajaknya sendiri sesuai dengan peraturan perundang-undangan perpajakan yang berlaku, maka kepatuhan WP menjadi aspek penting. Salah satu upaya yang dapat dilakukan oleh Direktorat Jenderal Pajak (DJP) sebagai otoritas pajak di Indonesia untuk meminimalisir resiko ketidakpatuhan WP adalah dengan melakukan pengawasan dan pemeriksaan terhadap WP (Darmayani & Herianti, 2017).

Sebagai penopang penerimaan terbesar, WP badan berdasarkan sektor usahanya mempunyai kontribusi dominan terhadap penerimaan negara setiap tahunnya. Banyaknya Klasifikasi Lapangan Usaha (KLU) atau jenis usaha WP, yang harus diawasi oleh pegawai pajak serta belum adanya sistem yang dapat membantu dalam mengetahui sektor yang potensial merupakan salah satu faktor yang mempengaruhi penerimaan pajak tidak mencapai target yang telah ditetapkan. Dengan adanya keterbatasan SDM membuat pengawasan dan pemeriksaan, khususnya terhadap WP badan, menjadi kurang optimal. Pengawasan dan pemeriksaan yang kurang optimal dapat berakibat tidak tertagihnya penerimaan pajak negara.

Penelitian sebelumnya membahas penggunaan algoritma K-Means dalam pengelompokan kunjungan wisatawan ke objek unggulan di provinsi DKI Jakarta (Maulida, 2018). Penelitian ini mengelompokkan berdasarkan tingkat kunjungan, yaitu tinggi, sedang, dan rendah. Penelitian lainnya adalah penggunaan algoritma K-Means dalam pengelompokan ayat Al Quran (Robani & Widodo, 2016). Penelitian dilakukan dengan terlebih dahulu melakukan prapemrosesan pada data teks dan melakukan pembobotan terhadap setiap kata yang dihasilkan.

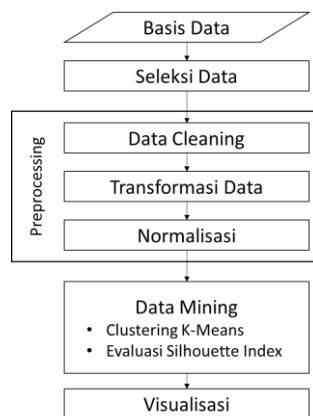
Karena latar belakang tersebut, peneliti mencoba menggunakan algoritma K-Means untuk mengelompokkan Klasifikasi Lapangan Usaha (KLU) atau jenis usaha WP yang memiliki potensi bagi penerimaan pajak dari data yang dimiliki oleh DJP. Kelebihan K-Means adalah mudah untuk diadaptasi dan waktu yang dibutuhkan untuk menjalankan relatif cepat. Selain itu untuk dataset yang besar, algoritma ini relatif sederhana, efisien dan mudah dipahami.

## 2. METODE PENELITIAN

Penelitian ini dimulai dengan identifikasi masalah yang dilakukan melalui pengamatan dan survey pada instansi terkait, dilanjutkan dengan pengumpulan data, pengolahan data dan pembahasan.

Pengolahan data menerapkan konsep *preprocessing* data sebelum data di *clustering*. Tujuan *preprocessing* adalah mentransformasi data ke suatu format yang menjadikan proses data mining lebih mudah dan efektif sesuai kebutuhan.

*Preprocessing* juga membantu dalam mendapatkan hasil yang lebih akurat, mengurangi waktu untuk penghitungan dalam skala besar, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi di dalamnya. Gambar 1 menunjukkan tahapan yang dilakukan dalam penelitian ini.



Gambar 1 Tahapan Penelitian

### 2.1 Seleksi data

Data yang akan digunakan dalam penelitian ini adalah data yang diperoleh dari Direktorat Data dan Informasi Perpajakan, Direktorat Jenderal Pajak, Kementerian Keuangan Republik Indonesia dalam rentang tahun 2016 s.d tahun 2019 yang mencakup sembilan Kantor Pelayanan Pajak (KPP). Data yang digunakan yaitu data Registrasi, data Tanda Terima, data Surat Pemberitahuan (SPT), data Riwayat Pemeriksaan, data Penerimaan, dan data Pengembalian. Kemudian dilakukan seleksi fitur data dengan memilih kolom-kolom yang diperlukan.

### 2.2 Data Cleaning

Pada tahap ini dilakukan pembersihan data yang mempunyai *missing value*, *redundant*, atau tidak relevan dengan penelitian. Data tersebut kemudian dihapus dari dataset. Jumlah data awal dan data setelah dilakukan *cleaning* dapat dilihat pada Tabel 1.

Tabel 1 Jumlah Data

No	Tabel	Jumlah Data	
		Data Awal	Data Akhir
1	Registrasi	5.127	4.101
2	Tanda Terima	4.618	4.618
3	SPT – Biaya Lain	4.594	3.656
4	Riwayat Pemeriksaan	2.520	2.389
5	Penerimaan	4.829	3.896
6	Pengembalian	2.242	1.902
7	SPT	4.619	3.756
<b>Total</b>		<b>28.549</b>	<b>24.318</b>

### 2.3 Transformasi Data

Langkah selanjutnya adalah transformasi data. Dilakukan proses skoring dari masing-masing variable agar dapat dilakukan *clustering*. Data dikelompokkan menjadi 2 kategori yakni variabel A untuk melihat tingkat ketidakpatuhan dan variabel B untuk melihat dampak fiskalnya terhadap

penerimaan. Variabel yang digunakan pada variabel A adalah sebagai berikut:

1. Pelaporan SPT tahunan badan sebagai variabel A1, variabel ini bertujuan untuk melihat kepatuhan formal WP badan yang tidak melaporkan SPT tahunannya dengan menyandingkan data tanda terima SPT tahunan dengan data registrasi WP badan per kelompok KLU.
2. Ekualisasi omset Pajak Penghasilan (PPh) dan penyerahan Pajak Pertambahan Nilai (PPN) sebagai variabel A2 yang diambil dari perbandingan nilai peredaran usaha dan nilai Dasar Pengenaan Pajak (DPP) PPN.
3. Ekualisasi perolehan PPN terhadap peredaran usaha PPh sebagai variabel A3 untuk melihat perbandingan antara pembelian dari data perolehan SPT Masa PPN terhadap peredaran usaha WP badan per kelompok KLU.
4. Histori pemeriksaan sebagai variabel A4 untuk menentukan WP badan per kelompok KLU yang dilakukan pemeriksaan dengan melihat apakah terdapat data ketetapan atas WP badan dalam kelompok KLU selama rentang waktu tertentu.
5. Rasio biaya lain-lain terhadap biaya total sebagai variabel A5 untuk mengidentifikasi WP Badan per kelompok KLU yang memperkecil keuntungan dengan melaporkan biaya lain-lain yang dinilai tidak wajar.

Selanjutnya masing-masing variabel A akan di skoring dengan nilai 0 – 100 berdasarkan hal-hal seperti pada Tabel 2. Semakin tinggi nilai skor menunjukkan semakin tinggi tingkat ketidakpatuhan.

Pada variabel B, variable yang digunakan seperti berikut:

1. Jumlah peredaran usaha sebagai variabel B1 untuk memperoleh nilai peredaran usaha WP badan per kelompok KLU.
2. Total biaya sebagai variabel B2 untuk memperoleh total biaya yang dilaporkan oleh WP badan per kelompok KLU.
3. Pengembalian pajak sebagai variabel B3 untuk memperoleh jumlah restitusi yang diterima oleh WP Badan per kelompok KLU.
4. Pajak yang disetor sebagai variabel B4 untuk memperoleh jumlah pembayaran pajak yang dilakukan WP badan per kelompok KLU.

Tabel 2 Skoring

Variabel	Keterangan	Skor
A1	Lapor SPT 4 tahun	0
	Tidak lapor 1 tahun	50
	Tidak lapor > 1 tahun	100
A2	-10% < Ekualisasi Omset < 10%	0
	Ekualisasi Omset < -10% or > 10%	80
	Ekualisasi Omset < -20% or > 20%	100
A3	Ekualisasi Perolehan ≤ 50%	0
	50% < Ekualisasi Perolehan < 100%	75

Variabel	Keterangan	Skor
A4	Ekualisasi Perolehan ≥ 100%	100
	Tidak ada ketetapan pajak	0
	Ketetapan pajak 1 Tahun	50
A5	Ketetapan pajak > 1 Tahun	100
	Rasio Total < 30%	0
	Rasio Total ≥ 30%	100

Tabel 3 Hasil Skoring Variabel A

idA	klu	kpp adm	A1	A2	A3	A4	A5
1	1262	92	0	0	43,75	100	0
2	1262	51	0	0	100	100	0
3	2117	92	100	90	0	100	0
4	2117	91	0	0	0	50	0
5	10320	92	0	20	93,75	100	0
...	...	...	...	...	...	...	...
1013	47511	56	0	0	0	0	0

Tabel 4 Hasil Skoring Variabel B

idB	klu	kpp adm	B1	B2	B3	B4
<i>(dalam 10.000.000.000)</i>						
1	1262	92	156	152	-5	1
2	1262	51	2.100	2.020	-6	64
3	1262	57	0	0	0	2
4	1262	58	0	0	-46	116
5	2117	92	232	219	0	10
...	...	...	...	...	...	...
1013	47511	56	0	0	0	0,01

### 2.4 Normalisasi

Dalam penelitian ini metode normalisasi yang digunakan adalah *z-score*. Metode *z-score* merupakan metode normalisasi berdasarkan nilai rata-rata dan deviasi standar dari populasi data (Caroline & Laturette, 2019; Nasution et al., 2019). Persamaan (1) merupakan persamaan yang digunakan dalam metode *z-score*.

$$v' = \frac{v - \mu}{\sigma} \tag{1}$$

Pada Persamaan (1), *v'* adalah nilai yang baru. *v* adalah nilai lama. *μ* merupakan rata-rata dari populasi. Sedangkan *σ* adalah standar deviasi dari populasi. Data normalisasi variabel A dan B masing-masing ditunjukkan pada Tabel 5 dan 6.

Tabel 5 Data Normalisasi Variabel A

idA	klu	kpp adm	A1	A2	A3	A4	A5
1	1262	92	-0,4	-0,5	0,7	1,5	-0,3
2	1262	51	-0,4	-0,5	2,3	1,5	-0,3
3	2117	92	-0,4	-0,5	-0,5	-0,7	-0,3
4	2117	91	-0,4	-0,5	-0,5	-0,7	-0,3
5	10320	92	2,9	2,3	-0,5	1,5	-0,3
...	...	...	...	...	...	...	...
1013	47511	56	-0,4	-0,5	-0,5	-0,7	-0,3

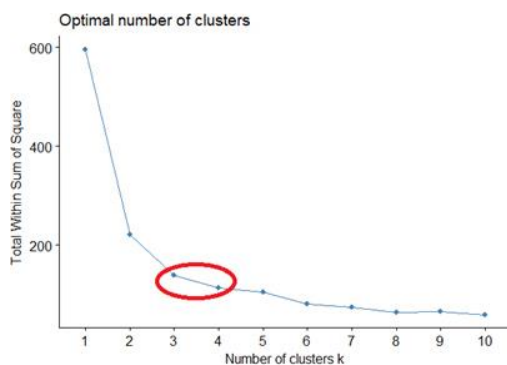
Tabel 6 Data Normalisasi Variabel B

idB	klu	kpp adm	B1	B2	B3	B4
1	1262	92	-0,1	-0,1	0	-0,2
2	1262	51	0,6	0,7	0	0,1
3	1262	57	-0,2	-0,2	0,2	-0,2
4	1262	58	-0,2	-0,2	-1,5	0,3
5	2117	92	-0,1	-0,1	0,2	-0,1
...	...	...	...	...	...	...
1013	47511	56	-0,2	-0,2	0,2	-0,2

## 2.5 Data Mining

Tahap ini merupakan proses menemukan informasi, pola, dan korelasi dalam dataset menggunakan teknik atau metode tertentu. Metode yang akan digunakan pada penelitian ini adalah metode *clustering* dengan algoritma K-Means.

K-Means merupakan salah satu algoritma yang berbasis titik pusat (*centroid*). Metode ini secara iteratif membagi kumpulan objek ke dalam *cluster* berbeda (Satria et al., 2019) yang telah ditentukan sebelumnya di mana setiap objek hanya dimiliki oleh satu *cluster* dan tidak tumpang tindih. K-Means membagi objek yang memiliki tingkat kemiripan maksimal dalam satu *cluster* dan memiliki kemiripan minimal dengan objek dari *cluster* lainnya (Li & Wu, 2012). Jika jumlah *cluster* belum ditentukan oleh pengguna, maka untuk menentukan jumlah *cluster* optimal dapat menggunakan metode *elbow*. Untuk menentukan nilai K metode *elbow* dapat dilakukan dengan melihat pada grafik yang pergerakannya mulai landai dari yang pertama setelah curam (Dewi & Pramita, 2019), seperti pada Gambar 2, dapat dilihat bahwa grafik yang pergerakannya landai pertama yaitu pada jumlah K *cluster* 3.



Gambar 2 Contoh Grafik Elbow

Tahapan yang dilakukan pada proses *clustering* K-Means, yaitu:

- Menentukan jumlah *cluster* K yang akan dibentuk
- Inisialisasi *centroid* (titik pusat *cluster*) awal sebanyak K *cluster*
- Menghitung jarak antara obyek dan pusat *cluster*
- Menempatkan setiap objek ke dalam kelompok *cluster* yang memiliki jarak terdekat dengan *centroid*

e. Menghitung kembali *centroid* yang terbentuk dengan mengambil rata-rata dari semua objek yang dimiliki setiap *cluster* sebagai posisi *centroid* baru

f. Mengulang sampai sampai tidak ada perpindahan objek antar *cluster*

Metode evaluasi yang akan digunakan pada penelitian ini adalah metode koefisien *silhouette* atau *silhouette index*. Metode ini mengevaluasi penempatan setiap objek dalam setiap *cluster* dengan membandingkan jarak rata-rata entitas di dalam satu *cluster* dan jarak antar entitas dalam *cluster* yang berbeda (Nur Aini et al., 2016). Dengan rentang hasil 1 dan -1, jika hasil mendekati 1 menunjukkan bahwa objek cocok dengan *cluster* nya sendiri dan tidak cocok dengan *cluster* tetangga (Nahdliyah et al., 2019). Semakin tinggi nilai koefisien *silhouette* maka semakin baik kualitas *cluster* yang terbentuk (Anggara et al., 2016; Gentle et al., 1991).

## 2.6 Visualisasi

Pola informasi yang dihasilkan dari tahap sebelumnya kemudian ditampilkan dalam bentuk kuadran agar lebih mudah dipahami. Visualisasi ini dapat digunakan untuk mengetahui WP badan yang perlu dilakukan pengawasan dan pemeriksaan lebih lanjut.

## 3. HASIL DAN PEMBAHASAN

*Clustering* dikelompokkan menjadi 2 kategori yakni tingkat ketidakpatuhan dan dampak fiskal yang disebabkan, masing-masing disimbolkan dengan variabel A dan variabel B. *Clustering* pada variabel A terdiri dari beberapa variabel antara lain, laporan spt tahunan, peredaran usaha, dan riwayat pemeriksaan. Sedangkan variabel yang digunakan untuk *clustering* variabel B adalah nilai peredaran usaha, total biaya yang dilaporkan, jumlah pengembalian, dan jumlah penerimaan. Selanjutnya KLU akan di *cluster* (kelompokkan) menjadi 3, yaitu risiko tinggi, sedang dan rendah. *Clustering* dilakukan menggunakan perangkat lunak R Studio version 1.4.1103.

### 3.1 Clustering Variabel A

Setelah data berhasil diambil dan disimpan ke dalam program R, kemudian ditentukan kolom yang akan digunakan dalam proses *clustering*. Pada *clustering* variabel A, kolom yang digunakan adalah kolom 4 sampai dengan 8 (lihat Tabel 2), maka kolom 1 (idA), 2 (klu), dan 3 (kppadm) dihapus dari data.

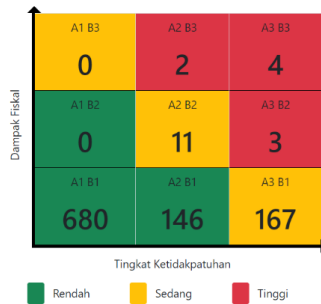
Selanjutnya *clustering* dilakukan dengan menggunakan metode K-Means, dengan jumlah pusat *cluster* k = 3. Pada tahap ini inisiasi atau titik pusat *cluster* dipilih secara acak oleh perangkat lunak R. Hasil *clustering* yang dihasilkan dapat dilihat pada Tabel 7 dimana kolom C menunjukkan *Cluster*.

Setiap data pada masing-masing kategori pada variabel A di *cluster* menjadi 3 kelompok yakni risiko tinggi, sedang dan rendah. Dari keluaran hasil K-









Gambar 8 Visualisasi Cluster Variabel A dan B

Penjelasan dari masing-masing bagian kuadran adalah sebagai berikut:

1. Kuadran A1-B1, berwarna hijau, terdapat 680 KLU dengan ketidapatuhan rendah dan dampak fiskal rendah.
2. Kuadran A1-B2, berwarna hijau, terdapat 0 KLU dengan ketidapatuhan rendah dan dampak fiskal sedang.
3. Kuadran A1-B3, berwarna kuning, terdapat 0 KLU dengan ketidapatuhan rendah dan dampak fiskal tinggi.
4. Kuadran A2-B1, berwarna hijau, terdapat 146 KLU dengan ketidapatuhan sedang dan dampak fiskal rendah.
5. Kuadran A2-B2, berwarna kuning, terdapat 11 KLU dengan ketidapatuhan sedang dan dampak fiskal sedang.
6. Kuadran A2-B3, berwarna merah, terdapat 2 KLU dengan ketidapatuhan sedang dan dampak fiskal tinggi.
7. Kuadran A3-B1, berwarna kuning, terdapat 167 KLU dengan ketidapatuhan tinggi dan dampak fiskal rendah.
8. Kuadran A3-B2, berwarna merah, terdapat 3 KLU dengan ketidapatuhan tinggi dan dampak fiskal sedang.
9. Kuadran A3-B3, berwarna merah, terdapat 4 KLU dengan ketidapatuhan tinggi dan dampak fiskal tinggi.

Hasil pengujian menunjukkan terdapat 9 KLU yang masuk ke dalam kuadran berwarna merah dengan tingkat ketidapatuhan tinggi (variabel A) dan memiliki dampak fiskal yang tinggi (variabel B), yaitu kuadran A2-B3, A3-B2, dan A3-B3 dengan masing-masing kuadran terdapat 2, 3, dan 4 KLU.

#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pengujian terhadap pengelompokan KLU menggunakan metode K-Means clustering dengan jumlah 3 pusat cluster dapat diambil kesimpulan sebagai berikut.

1. Pada variabel A kelompok yang memiliki nilai rata-rata cluster tinggi dianggap memiliki tingkat ketidapatuhan tinggi. Begitu juga pada variabel B, kelompok yang memiliki nilai rata-rata cluster yang tinggi, dianggap kelompok yang memiliki dampak fiskal tinggi terhadap penerimaan. Pengawasan dan pemeriksaan dapat dilakukan

dengan memprioritaskan 9 KLU yang masuk ke dalam kuadran berwarna merah dengan tingkat ketidapatuhan tinggi (variabel A) dan memiliki dampak fiskal yang tinggi (variabel B). Dilanjutkan dengan kuadran berwarna kuning dan hijau.

2. Nilai rata-rata koefisien *silhouette* yang didapatkan untuk variabel A dan variabel B masing-masing adalah 0,65 dan 0,93. Hasil *clustering* dikatakan baik jika nilai koefisien *silhouette* bernilai positif dan mendekati nilai 1. Sehingga kualitas *cluster* yang terbentuk pada penelitian ini dapat dikatakan baik.

Adapun saran yang bisa diberikan adalah sebagai berikut.

1. Perlu diterapkan metode *preprocessing* yang lebih baik agar hasil *cluster* yang terbentuk lebih akurat.
2. Selain menggunakan data internal dari DJP sebaiknya juga dilengkapi dengan data eksternal seperti dari instansi, lembaga, asosiasi, dan pihak lain (ILAP).

#### DAFTAR PUSTAKA

- ANGGARA, M., SUJAINI, H., & NASUTION, H. 2016. Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 4(1), 186–191.
- CAROLINE, L., & LATURETTE, K. 2019. Comparative Analysis of Z-Score and Springate Altman Models on Registered Coal Companies Bei in. *Research In Management and Accounting*, 2(2), 56–66.
- DARMAYANI, D., & HERIANTI, E. 2017. Pengaruh Tingkat Kepatuhan Wajib Pajak Badan Terhadap Peningkatan Penerimaan Pajak Penghasilan Dengan Penagihan Pajak Sebagai Variabel Moderating (Pada KPP Pratama Cilandak Jakarta Selatan). *InFestasi*, 13(1), 275.
- DEWI, D. A. I. C., & PRAMITA, D. A. K. 2019. Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix : Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109.
- GENTLE, J. E., KAUFMAN, L., & ROUSSEUW, P. J. 1991. Finding Groups in Data: An Introduction to Cluster Analysis. In *Biometrics* (Vol. 47, Issue 2).
- HANIFAH, I. S., & AGUNG S, R. E. W. 2016. Pengaruh Kepatuhan Wajib Pajak Dan Penagihan Pajak Terhadap Penerimaan Pph Pasal 25/29 Wajib Pajak Badan Pada Kpp Pratama Batang. *Maksimum*, 3(1), 1.
- LI, Y., & WU, H. 2012. A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109.
- MAULIDA, L. 2018. Penerapan Datamining Dalam

- Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167.
- NAHDLIYAH, M. A., WIDIHARIH, T., & PRAHUTAMA, A. 2019. Metode K-Medoids Clustering Dengan Validasi Silhouette Index Dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian*, 8(2), 161–170.
- NASUTION, D. A., KHOTIMAH, H. H., & CHAMIDAH, N. 2019. Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78.
- NUR AINI, F., PALGUNADI, S., & ANGGRAININGSIH, R. 2016. Clustering Business Process Model Petri Net Dengan Complete Linkage. *Jurnal Teknologi & Informasi ITSmart*, 3(2), 47.
- RACZKOWSKI, K., & MRÓZ, B. 2018. Tax gap in the global economy. *Journal of Money Laundering Control*, 21(4), 567–583.
- ROBANI, M., & WIDODO, A. 2016. Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia. *Jurnal Sistem Informasi Bisnis*, 6(2), 164.
- SATRIA, E., TAMBUNAN, H. S., SARAGIH, I. S., DAMANIK, I. S., & SITANGGANG, F. T. E. 2019. Penerapan Clustering dalam Mengelompokkan Jumlah Kunjungan Wisatawan Mancanegara Dengan Metode K-Means. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 462.
- SUBROTO, G. 2020. *Artikel - MEMAHAMI TAX GAP*. 9 Januari. [online] Tersedia di: <https://bppk.kemenkeu.go.id/content/artikel/balai-diklat-keuangan-denpasar-memahami-tax-gap-2020-01-09-6bfb976f> [Diakses 10 Desember 2020]