

ANALISIS *CLUSTERING* DAN KLASIFIKASI SATKER K/L BERDASARKAN NILAI KINERJA PELAKSANAAN ANGGARAN MENGUNAKAN *MACHINE LEARNING*

Catur Ery Prabowo, Sabar Sautomo, Muhammad Fajri Natsir, Legendani, Putri Alfafa
Direktorat Jenderal Perbendaharaan, Kementerian Keuangan, Jakarta

ABSTRACT

Monitoring and evaluation (monev) of budget execution is part of the process to support better management of government spending, including aspects of monitoring and evaluation of budget execution, including performance of budget execution in the units (satker) of Ministries/Agencies. Utilization of budget execution performance data with clustering and classification methods can be used to get an overview of the types and characteristics of the Ministry/Agency units, so that the monitoring and evaluation process for budget execution is more effective and efficient. The Clustering method in this study uses the K-Means algorithm, then combined with the Classification method of several Machine Learning algorithms, including K-Nearest Neighbor (K-NN), Logistic Regression (LR), Classification And Regression Trees (CART) and Random Forest (RF), from the evaluation results of the classification algorithm that the Random Forest (RF) algorithm produces a better accuracy value than the other algorithms, which is 0.985

Keywords: Kinerja Pelaksanaan Anggaran, K-Mean, Clustering, Machine Learning

ABSTRAK

*Monitoring dan evaluasi (monev) pelaksanaan anggaran adalah bagian dari proses untuk mendukung pengelolaan belanja pemerintah yang lebih baik, diantara aspek monev pelaksanaan anggaran antara lain Kinerja pelaksanaan anggaran pada satuan kerja (satker) Kementerian/Lembaga. Pemanfaatan data-data kinerja pelaksanaan anggaran dengan metode *clustering* dan klasifikasi dapat digunakan untuk mendapatkan gambaran dari tipe dan karakteristik satker Kementerian/Lembaga, sehingga proses monev pelaksanaan anggaran lebih efektif dan efisien. Metode *Clustering* pada penelitian ini menggunakan algoritma K-Means, kemudian dikombinasikan dengan metode Klasifikasi dari beberapa algoritma *Machine Learning*, antara lain K-Nearest Neighbor (K-NN), Logistic Regression (LR), Classification And Regression Trees (CART) dan Random Forest (RF), dari hasil evaluasi algoritma klasifikasi tersebut bahwa algoritma Random Forest (RF) menghasilkan nilai akurasi lebih baik dari algoritma yang lain yaitu sebesar 0.985*

Kata kunci: Kinerja Pelaksanaan Anggaran, K-Mean, Clustering, Machine Learning

PENDAHULUAN

Latar Belakang

Belanja Kementerian Negara/Lembaga (K/L) sebagai bagian dari Belanja Negara dalam APBN dipahami menjadi instrumen penting kebijakan fiskal pemerintah dalam rangka pembangunan dan pertumbuhan ekonomi. Bahkan di tahun 2020 – 2021 terlihat bagaimana Belanja K/L bekerja keras menjadi salah satu *backbone* upaya pemerintah menangani pandemi Covid-19 dan pemulihan ekonomi nasional (Kemenkeu RI, 2021).

Mengingat pentingnya peran Belanja K/L, pemerintah senantiasa melakukan langkah-langkah perbaikan pengelolaan dalam rangka mewujudkan belanja yang berkualitas. Salah satu langkahnya adalah melalui kebijakan monitoring dan evaluasi (monev) pelaksanaan anggaran belanja K/L sebagaimana diatur dalam Peraturan Menteri Keuangan (PMK) Nomor 195/PMK.05/2018. Monev dilakukan untuk menjamin efektivitas dan efisiensi pelaksanaan anggaran dengan tetap patuh dan taat terhadap regulasi pelaksanaan anggaran (Peraturan menteri keuangan Republik Indonesia, 2018).

Perbaikan pengelolaan belanja juga dilakukan melalui pemanfaatan teknologi informasi dan digitalisasi yang bertujuan untuk mempengaruhi beberapa aspek dan memberikan perubahan yang signifikan, baik dari sisi pemerintah sebagai pengelola keuangan negara, maupun pada *stakeholders* yang menjadi pengguna layanan keuangan negara (Penelitian & Kelembagaan, 2017)

Selain aspek proses operasional pelaksanaan anggaran, pemanfaatan teknologi informasi juga diterapkan di aspek monev pelaksanaan anggaran. Tools monev berupa Indikator Kinerja Pelaksanaan Anggaran (IKPA) dilakukan dengan pengukuran pada setiap indikator IKPA secara digital pada sistem aplikasi OMSPAN Direktorat Jenderal Perbendaharaan Kementerian Keuangan sejak tahun 2018. Pengukuran IKPA diterapkan pada seluruh Satker K/L, sehingga menghasilkan data dan informasi nilai setiap indikator IKPA sebagai representasi kualitas kinerja pelaksanaan anggaran setiap Satker.

Tantangan saat ini adalah tentang bagaimana memanfaatkan data yang tersedia tersebut untuk optimalisasi pelaksanaan pembinaan Satker dalam kerangka monev pelaksanaan anggaran.

Mengingat beragamnya Satker pelaksana anggaran Belanja K/L dengan karakteristiknya masing-masing, maka model pembinaan dan monev selama ini yang mengandalkan intuisi dan penilaian subyektif para pembinaanya perlu diubah menjadi lebih baik. Idealnya, dengan adanya data kinerja IKPA setiap Satker tersebut, maka dapat menjadi pintu masuk untuk melihat profil dan karakteristik setiap Satker sebagai perwujudan "*know your customer*".

Pada hasil penelitian yang dilakukan (Małeck & Watróbski, 2017) bahwa dengan melakukan analisis terhadap *clustering* dan klasifikasi tipe konsumen merupakan kunci sukses dalam melakukan teknik pemasaran efektif yang dilakukana oleh perusahaan.

Pada penelitian ini, penulis akan melakukan eksplorasi terhadap data-data indikator kinerja pelaksanaan anggaran untuk melihat keterkaitan dan hubungan antar indikator kinerja pelaksanaan anggaran, kemudian dilakukan *clustering* dan klasifikasi untuk mendapatkan gambaran mengenai *cluster* atau kelompok Satker K/L berdasarkan kemiripan karakteristik nilai indikator IKPA. *Cluster* Satker K/L tersebut dapat menjadi *insight* bagi pemerintah dalam rangka optimalisasi monev dan pembinaan pelaksanaan anggaran karena kegiatan lebih tepat sasaran terhadap setiap Satker K/L sesuai kebutuhan dan karakteristiknya masing-masing.

TINJAUAN LITERATUR

Indikator Kinerja Pelaksanaan Anggaran

Indikator Kinerja Pelaksanaan merupakan indikator yang ditetapkan oleh Kementerian Keuangan selaku BUN untuk mengukur kualitas kinerja pelaksanaan anggaran belanja Kementerian Negara/Lembaga dari sisi kesesuaian terhadap perencanaan, efektivitas pelaksanaan anggaran, efisiensi pelaksanaan anggaran, dan kepatuhan terhadap regulasi (Peraturan Ditjen Perbendaharaan, 2021).

Clustering Dataset

Algoritma K-means adalah algoritma pengelompokan (*Clustering*) tipe data numerik yang paling populer dikembangkan hingga saat ini, karena efisiensi dan skalabilitasnya yang tinggi

saat proses clustering pada data yang besar (Ben Salem et al., 2018)

Segementasi *stakeholders* bukan hanya untuk mendapatkan pengelompokan keragaman dataset dalam kelompok yang lebih kecil namun dapat digunakan untuk analisis dalam mendapatkan informasi yang lebih dari perilaku *stakeholders* (Yotsawat & Srivihok, 2016)

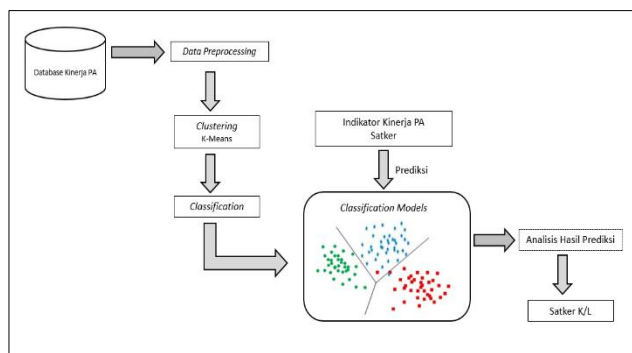
Untuk mendapatkan jumlah Cluster yang terbaik pada data yang besar digunakan teknik kombinasi metode K-Means dengan Elbow, dan teknik ini dapat meningkatkan efisiensi dan kinerja k-means yang efektif (Syakur et al., 2018)

Studi tentang data mining sangat bergantung pada kinerja *clustering*. *Clustering* sebelum klasifikasi disebut sebagai *cluster Classifier*. Dataset yang sebelumnya dilakukan proses cluster, kemudian dilakukan proses pemodelan klasifikasi, beberapa pemodelan klasifikasi antara lain K-Nearest Neighbor (K-NN), Logistic Regression (LR), Classification and Regression Trees (CART) dan Random Forest (RF) untuk mendapatkan hasil performa yang terbaik (Arumugam & Christy, 2018).

METODOLOGI PENELITIAN

Dalam implementasi metodologi yang digunakan seperti pada Gambar 1 Kerangka Pemikiran, penulis melakukan eksperimen dengan menggunakan *tools library Python*, dari tahap Pemrosesan data (*Data Preprocessing*), *Clustering*, *Classification* hingga proses pemodelan dari algoritma yang digunakan.

Gambar 1. Kerangka Pemikiran



Sumber: Penulis

Tahap pertama yang dilakukan adalah pemrosesan data yang akan digunakan untuk

permodelan. Pada tahap ini, *missing value* pada dataset dibersihkan sehingga data lebih mudah diidentifikasi dan dapat meningkatkan proses pembelajaran dari algoritma *machine learning* yang digunakan.

Metode cluster data yang digunakan dalam penelitian ini adalah metode K-Means Clustering. Metode K-Means Clustering adalah metode yang menggunakan *descriptive model* untuk mengungkapkan algoritma pengelompokan variabel dengan karakteristik yang sama.

Penelitian ini mengkombinasikan metode K-Means dengan metode elbow untuk mengetahui jumlah cluster optimal satuan kerja berdasarkan nilai masing-masing IKPA. Nilai pada k akan terus meningkat pada setiap proses dan berkurang dengan nilai yang besar. Penelitian ini mencari nilai k terbaik dengan menggunakan metode elbow. Metode elbow diterapkan dengan melihat posisi siku dalam inersia dan nilai k ideal. Hasil cluster k terbaik akan menjadi dasar untuk clustering.

Proses dilanjutkan dengan mengambil data x dari dataset sebagai sentroid awal yang mewakili sebuah cluster. Menurut metode K-Means Clustering, pengelompokan bergantung pada pilihan sentroid awal. Perhitungan pada metode K-Means Clustering dilakukan dengan rumus Euclidian terhadap sentroid. Elemen data yang memiliki jarak paling kecil ke sentroid dipindahkan ke cluster yang sesuai. Proses tersebut dilakukan berulang sampai tidak ada perubahan lagi pada cluster.

Pada penelitian ini satuan kerja akan dikelompokkan berdasarkan kecenderungan persamaan nilai. Pada tahap ini nilai IKPA satuan kerja yang telah diambil dan dilakukan data cleaning akan dikelompokkan sesuai dengan kemiripan karakteristik nilai.

Informasi hasil permodelan kemudian dianalisis secara deskriptif dengan didukung oleh beberapa teknik visualisasi data sebagai acuan pengambilan keputusan.

HASIL DAN PEMBAHASAN

Eksplorasi Dataset

Dataset yang digunakan bersumber dari data kinerja pelaksanaan anggaran satker Kementerian/Lembaga pada database Sistem

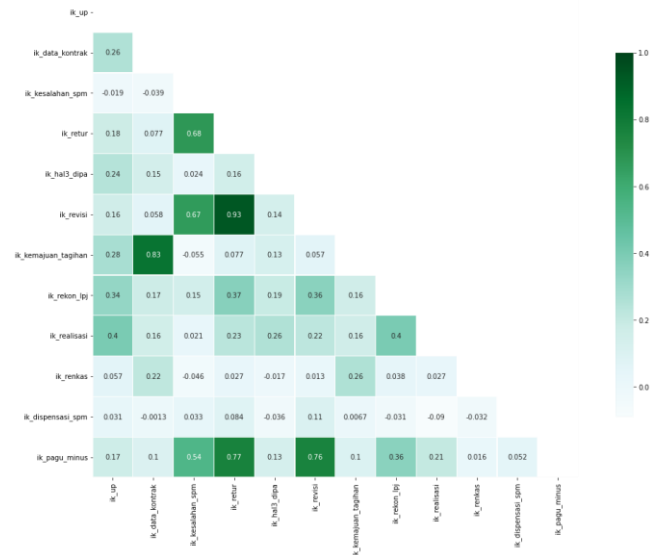
Perbendaharaan dan Anggaran Negara (SPAN) dari tahun 2018 sampai dengan tahun 2021, jumlah *record* data sebanyak 944,157. Variabel yang digunakan dalam dataset dapat dilihat pada Tabel 1.

Tabel 1. *Variables* Dataset

No	Variables	Keterangan
1	Tahun	Tahun anggaran
2	Bulan	Periode bulan
3	kodesatker	Kode satker
4	ik_up	Kinerja Uang Persediaan
5	ik_data_kontrak	Kinerja kontrak
6	ik_kesalahan_spm	Kinerja kesalahan SPM
7	ik_retur	Kinerja Retur SP2D
8	ik_hal3_dipa	Kinerja Hal 3 DIPA
9	ik_revisi	Kinerja Revisi DIPA
10	ik_kemajuan_tagihan	Kinerja kemajuan tagihan
11	ik_rekon_lpj	Kinerja laporan LPJ
12	ik_realisasi	Kinerja realisasi belanja
13	ik_renkas	Kinerja rencana kas
14	ik_dispensasi_spm	Kinerja dispensasi SPM
15	ik_pagu_minus	Kinerja Pagu Minus

Dari variabel yang sudah ditentukan, kemudian disusun matriks korelasi untuk menunjukkan keterkaitan antar atribut. Setiap atribut acak (X_i) dalam tabel berkorelasi dengan masing-masing atribut lain dalam tabel (X_j). Nilai korelasi berkisar antara -1 sampai dengan 1. Nilai korelasi yang mendekati 1 menunjukkan bahwa kedua atribut memiliki korelasi atau keterkaitan yang tinggi. Hasil *correlation matrix* indikator kinerja pada satuan kerja dapat dilihat pada Gambar 2 di bawah ini, dimana diperoleh pengetahuan bahwa nilai tertinggi adalah 0,93 yaitu korelasi antara kinerja revisi DIPA (ik_revisi) dan kinerja retur SP2D (ik_retur). Angka korelasi kedua variabel tersebut menunjukkan bahwa semakin banyak revisi dilakukan oleh Satker maka berpotensi meningkatkan volume retur SP2D, karena rentan terjadi ketidaksesuaian antara segmen anggaran yang terlanjur dicantumkan dalam dokumen SP2D dengan yang tertera pada DIPA setelah revisinya disahkan.

Gambar 2. *Correlation Matrix* Dataset

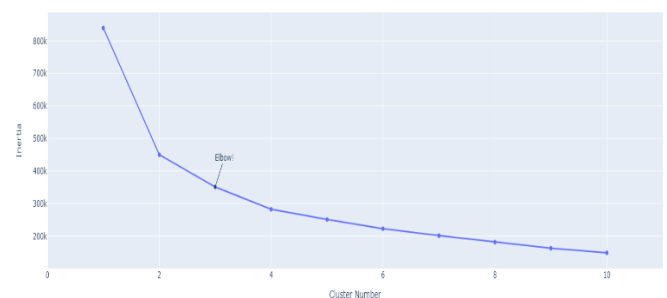


Penentuan Jumlah Cluster

Inersia mengukur seberapa baik kumpulan data dikelompokkan oleh K-Means. Inersia dihitung dengan mengukur jarak antara setiap titik data dan pusat massanya, mengkuadratkan jarak, dan menjumlahkan kuadrat di satu cluster.

Model yang baik adalah model dengan inersia rendah dan jumlah cluster (K) rendah. Namun, hal ini menjadi *tradeoff* karena ketika K meningkat, inersia menurun.

Gambar 3. Jumlah Cluster pada teknik Elbow



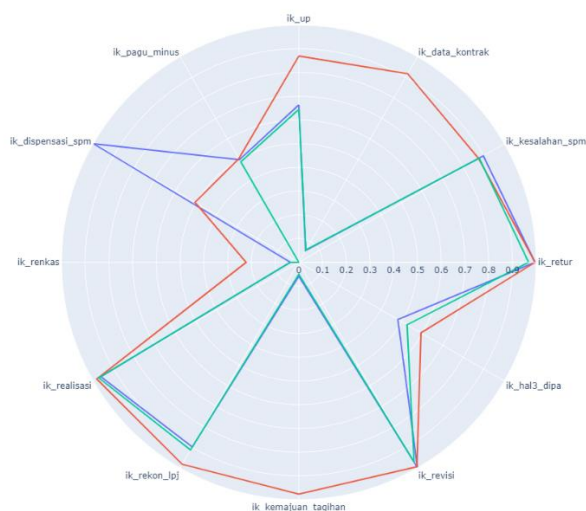
Metode elbow digunakan untuk menemukan K optimal pada dataset. Metode elbow menemukan titik di mana penurunan inersia mulai

melambat. Dalam penelitian ini, $K=3$ adalah “siku” seperti yang terlihat dalam grafik di atas. Jadi, jumlah cluster Satker optimal yang didapat dari metode elbow adalah sebanyak 3 cluster.

Proses Cluster

Hasil proses clustering menggunakan algoritma K-Means disajikan menggunakan grafik line_polar sebagaimana pada Gambar 4 berikut ini. Pemilihan grafik tersebut mengingat *insight* yang diperoleh dari analisis *clustering* akan dimanfaatkan kepada pihak bisnis (pemilik proses bisnis), sehingga dipilih penyajian yang relatif mudah dipahami (Mauricio Letelier, 2020).

Gambar 4. Hasil K-Means Clustering



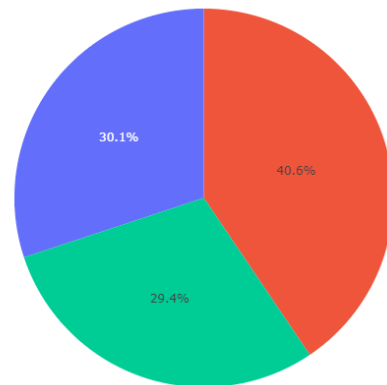
Pada Gambar 4 dapat dilihat 3 garis berwarna merah, biru dan hijau yang selanjutnya diberi label 0 untuk garis merah, label 1 garis biru dan label 2 untuk garis hijau. Dalam penelitian ini kami fokus pada variabel-variabel di setiap label yang paling berbeda dibanding di label lainnya. Sementara variabel yang relatif serupa seperti *ik_revisi* dan *ik_realisasi* tidak mendapat porsi dominan untuk diteliti lebih lanjut.

Pengetahuan yang diperoleh adalah bahwa label 0 dominan bernilai paling tinggi, terutama pada *ik_kemajuan_tagihan*, *ik_data_kontrak*, dan *ik_up*. Kemudian label 1 (garis biru) memiliki nilai tertinggi pada variabel *ik_dispensasi_SPM*, sedangkan sebaliknya label 2 memiliki nilai terendah untuk variabel *ik_dispensasi_SPM*.

Selain itu, sangat relevan untuk diketahui mengenai komposisi proporsi masing-masing label 0, label 1 dan label 2. Pada Gambar 5 di

bawah ini, ditunjukkan bahwa proporsi terbanyak adalah cluster label 0 yaitu mencapai 40,6% dari total, sedangkan terendah adalah label 2 yang sebesar 29,4% meskipun relatif sedikit bedanya dengan label 1 yang sebesar 30,1%.

Gambar 5. Proporsi K-Means Clustering



Berdasarkan analisis clustering di atas, untuk memudahkan pemahaman pemilik proses bisnis maupun pembaca, maka dibuatkan deskripsi untuk setiap cluster label 0, label 1 dan label 2 berdasarkan perilaku di setiap variabel indikator IKPA.

Untuk Label 0 karena dominan hampir di semua variabel mencatat nilai paling tinggi maka dapat dideskripsikan sebagai cluster Satker Disiplin Melaksanakan Anggaran. Label 1 adalah Satker Pengelola SPM Terbaik karena memiliki nilai tertinggi untuk *ik_dispensasi_SPM* dan *ik_kesalahan_SPM*. Sedangkan untuk Label 2 adalah Satker Perlu Perhatian karena beberapa variabelnya bernilai paling rendah terutama *ik_dispensasi_SPM*.

Gambar 6. Deskripsi Cluster



Klasifikasi Models

Model cluster satuan kerja ditentukan oleh hubungan masing-masing nilai indikator kinerja. Pemanku kepentingan dapat memahami basis

segmen yang secara akurat dapat memprediksi karakteristik satuan kerja dan menggunakan hasilnya untuk meningkatkan kinerja pelaksanaan anggaran. Bagian ini menyajikan model klasifikasi untuk memprediksi karakteristik satuan kerja sebagai bagian dari cluster yang sudah ada. Terdapat lima algoritma klasifikasi yang diterapkan dalam penelitian ini. Hasil penelitian menunjukkan bahwa algoritma Random Forest (RF) lebih baik dari algoritma yang lain dalam mengklasifikasikan satuan kerja berdasarkan indikator kinerja dengan nilai akurasi sebesar 0.985 seperti terlihat pada tabel 2 di bawah.

Tabel 2. Hasil Evaluasi *Classification Models*

Algoritma	Accuracy	precision	recall	f1-score
K-NN	0.971	0.970	0.967	0.971
LR	0.940	0.939	0.940	0.945
CART	0.956	0.955	0.952	0.956
RF	0.985	0.983	0.985	0.984
NB	0.970	0.965	0.966	0.967

KESIMPULAN

Hasil analisis *clustering* dan klasifikasi data indikator IKPA menggunakan metode *machine learning* menghasilkan informasi berupa adanya korelasi yang kuat antara kinerja revisi DIPA (ik_revisi) dengan kinerja retur SP2D (ik_retur). Kemudian berdasarkan data 12 indikator IKPA diperoleh *insight* menarik berupa 3 cluster Satker menurut karakteristik perilaku dalam pelaksanaan anggaran belanjanya, yaitu Satker Disiplin (Label 0), Satker Pengelola SPM Terbaik (Label 1), dan Satker Perlu Perhatian (Label 2).

Dengan model klasifikasi menggunakan algoritma Random Forest maka selanjutnya dapat diprediksi suatu Satker akan masuk kelompok label yang mana berdasarkan data indikator IKPA yang masuk ke dataset nantinya. Pengetahuan tersebut dapat digunakan untuk optimalisasi monev agar dilaksanakan lebih efisien dan efektif karena memiliki target Satker binaan yang lebih jelas dan spesifik. Dengan pelaksanaan monev dan pembinaan yang lebih terukur maka dapat mendukung pelaksanaan anggaran yang lebih baik untuk mewujudkan belanja pemerintah yang berkualitas.

REFERENSI

- Arumugam, P., & Christy, V. (2018). Analysis of Clustering and Classification Methods for Actionable Knowledge. *Materials Today: Proceedings*, 5(1, Part 1), 1839–1845. <https://doi.org/https://doi.org/10.1016/j.matpr.2017.11.283>
- Ben Salem, S., Naouali, S., & Chtourou, Z. (2018). A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. *Computers & Electrical Engineering*, 68, 463–483. <https://doi.org/https://doi.org/10.1016/j.compeleceng.2018.04.023>
- Kemenkeu RI. (2021). Kerja Keras APBN melalui Belanja Negara Berlanjut Hingga November 2021. <https://www.kemenkeu.go.id/Publikasi/Berita/Kerja-Keras-Apbn-Melalui-Belanja-Negara-Berlanjut-Hingga-November-2021>. <https://www.kemenkeu.go.id/publikasi/berita/kerja-keras-apbn-melalui-belanja-negara-berlanjut-hingga-november-2021>
- Małeck, K., & Watróbski, J. (2017). The Classification of Internet Shop Customers based on the Cluster Analysis and Graph Cellular Automata. *Procedia Computer Science*, 112, 2280–2289. <https://doi.org/10.1016/j.procs.2017.08.150>
- Mauricio Letelier. (2020). *Clustering With More Than Two Features? Try This To Explain Your Findings*. <https://towardsdatascience.com/clustering-with-more-than-two-features-try-this-to-explain-your-findings-b053007d680a>
- Penelitian, S., & Kelembagaan, K. (2017). *Treasury Policy Brief. September*, 1–5.
- Peraturan Ditjen Perbendaharaan. (2021). *Peraturan Ditjen Perbendaharaan Nomor PER- 4/PB/2021*.
- Peraturan menteri keuangan Republik Indonesia. (2018). *Monitoring Dan Evaluasi Pelaksanaan Anggaran Belanja Kementerian Negara/ Lembaga*. 195.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for

Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1). <https://doi.org/10.1088/1757-899X/336/1/012017>

Yotsawat, W., & Srivihok, A. (2016). Thai domestic tourists clustering model using machine learning Techniques: Case study of phranakhon si ayutthaya province, Thailand. *Information (Japan)*, 19(2), 413–422.

