

Penerapan Algoritma *K-Means* Untuk Mengetahui Tingkat Kepatuhan Wajib Pajak Kendaraan Bermotor Pada UPT Samsat Medan Selatan

Application of The K-Means Algorithm to Determine the Level of Motor Vehicle Tax Payer Compliance at UPT Samsat Medan Selatan

Dini Asti¹, Muhammad Siddik Hasibuan², Putri Aprilia Siregar³

^{1,2,3}Program Studi Ilmu Komputer Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

^{1,2,3}Jalan Lap.Golf No.120, Kp.Tengah, Kec. Pancur Batu, Sumatera Utara, Indonesia

E-mail: ¹diniasti888@gmail.com, ²muhammadsiddik@uinsu.ac.id,

³putriapriliasiregar10@gmail.com

Abstrak

Kepatuhan Wajib Pajak adalah tindakan menunjukkan patuh dan tertib terhadap kewajiban perpajakan dengan melakukan pembayaran pajak dan melaporkan pajak secara berkala oleh Wajib Pajak yang bersangkutan sesuai ketentuan perpajakan yang berlaku. Kelompok kepatuhan pajak kendaraan bermotor di samsat Medan Selatan dibagi menjadi banyak tingkatan dari rendah ke tinggi. Besaran iuran pajak kendaraan bermotor tergantung dari perhitungan dan pembayaran pajak terutang atas penghasilan yang diperoleh wajib pajak dan Pembayaran tunggakan pajak sebelum jatuh tempo. Tujuan penelitian ini adalah untuk mengetahui tingkat kepatuhan pajak kendaraan bermotor di samsat Medan Selatan tahun 2023. Algoritma K-Means, yang digunakan melalui metode clustering, menggunakan tahapan KDD, yang mencakup 146 data, yang berasal dari data pembayaran pajak kendaraan bermotor di Samsat Medan Selatan pada tahun 2023. Nilai determinasi cluster sebesar 0,294 dihasilkan oleh hasil pengujian RapidMiner yang menggunakan perhitungan indeks Davies-Bouldin. Dalam cluster 0 ada sepuluh wajib pajak dengan kepatuhan tingkat sangat rendah, di cluster 1 ada 56 wajib pajak dengan kepatuhan tingkat sedang, di cluster 2 ada 19 wajib pajak dengan kepatuhan tingkat rendah; dan di cluster 3, ada 61 wajib pajak dengan kepatuhan tingkat tinggi.

Kata Kunci : *Clustering; Data Mining; K-Means; Kepatuhan Wajib Pajak;*

Abstract

Taxpayer compliance is an action that shows obedience and order to tax obligations by making tax payments and reporting taxes periodically by the relevant taxpayer in accordance with applicable tax regulations. The group of motorized vehicle presence at the South Medan Samsat is divided into several levels, from lowest to highest. The amount of motor vehicle tax contributions depends on the calculation and payment of tax owed on the income earned by the taxpayer and payment of tax arrears before autumn. The aim of this research is to determine the level of motor vehicle tax compliance in South Medan Samsat in 2023. The research method used is the K-Means algorithm with the Clustering method. The K-Means algorithm uses the KDD stage with a total of 146 data in the form of motor vehicle tax payment data at the South Medan samsat in 2023. RapidMiner test results using the Davies-Bouldin Index calculation give a cluster determination value of 0.294. Cluster 0 contains 10 taxpayers with very low levels of compliance, cluster 1 contains 56 taxpayers with medium levels of compliance, cluster 2 contains 19 taxpayers with low levels of compliance, cluster 3 contains 61 taxpayers with high levels of compliance.,

Keywords: Clustering; Data Mining; K-Means; Taxpayer Compliance

1. PENDAHULUAN

Pajak merupakan sumber utama penerimaan negara, khususnya pada anggaran pendapatan dan peruntukan negara (APBN). Hasilnya digunakan untuk membiayai pengeluaran pemerintah dalam penyelenggaraan pemerintahan yang bertujuan untuk meningkatkan kesejahteraan rakyat melalui pembangunan dan peningkatan fasilitas umum. [1]

Kesadaran wajib pajak dibentuk oleh nilai-nilai moral, etika, dan rasa tanggung jawab sebagai warga negara yang sesungguhnya. Rendahnya kepatuhan wajib pajak di kalangan wajib pajak tercermin dari beberapa faktor, antara lain tingkat pendapatan, pemahaman peraturan perpajakan, dan kualitas pelayanan. Pendapatan wajib pajak yang berbeda mempunyai dampak yang berbeda pula terhadap keakuratan dan kepatuhan perpajakan. [2]

Tentu saja, seiring dengan meningkatnya perkembangan motor setiap tahunnya, Kantor Samsat akan memanfaatkannya untuk memungut pajak dari pemilik mobil guna meningkatkan kepatuhan wajib pajak mobil dan meningkatkan sumber daya keuangan asli daerah. Namun banyak wajib pajak yang tidak membayar pajak kendaraan tepat waktu. Hal ini dapat merugikan pemerintah daerah dan pembayar pajak. Oleh karena itu, perlu adanya peningkatan tingkat kepatuhan wajib pajak kendaraan bermotor. Hal ini dilakukan untuk meningkatkan kemampuan sistem dalam mengorganisir informasi pembayaran pajak kendaraan dan mengidentifikasi kewajiban pembayaran pajak kendaraan. Penelitian ini fokus pada pemanfaatan *data mining* untuk mengetahui kepatuhan pajak kendaraan. Hal ini dilakukan dengan menggunakan algoritma *K-Means clustering* yang membantu pemerintah mengetahui tingkat kepatuhan wajib pajak dalam membayar pajak kendaraan.

Irwan, Sahlan Sidjara, dan Asmelia Putri Aryati telah melakukan penelitian sebelumnya dengan judul "Pengelompokan Jenis Penerimaan Pajak di Kota Makassar Menggunakan Fuzzy Clustering". Tujuan dari penelitian sebelumnya adalah untuk mengklasifikasikan jenis pemungutan pajak yang dilakukan di Kota Makassar. Teknik *fuzzy clustering* digunakan dalam penelitian ini. Penelitian ini akan membantu pemerintah menilai kesejahteraan masyarakat dengan mengetahui kelompok mana yang menerima pajak tertinggi di Kota Makassar. [3]

Tujuan dari penelitian ini adalah untuk mensintesis tingkat kepatuhan pajak wajib pajak berdasarkan tingkat pengetahuannya khususnya mengenai pembayaran pajak kendaraan bermotor yang masih rendah. Penelitian ini mengklasifikasikan tingkat kepatuhan pajak wajib pajak menjadi beberapa kelompok yaitu kelompok kepatuhan tinggi, kelompok kepatuhan sedang, kelompok kepatuhan rendah, dan kelompok kepatuhan sangat rendah. Data pembayaran pajak kendaraan Samsat Medan Selatan diolah menggunakan *software* Rapidminer dengan teknik clustering dan algoritma K-Means.

Data mining adalah ilmu yang melibatkan penemuan pola-pola menarik dalam sejumlah besar data. Data dapat dimasukkan ke dalam database, spreadsheet, atau perangkat penyimpanan informasi lainnya. Penambangan data berhubungan dengan bidang lain seperti sistem basis data, pergudangan data, statistik, pembelajaran mesin, pengambilan informasi, komputasi tingkat lanjut, dan pengenalan bentuk. Penambangan data juga terkait dengan banyak bidang ilmu lainnya, seperti analisis data spasial, jaringan saraf, dan pengenalan pola. [4]

Penambangan data adalah teknik penambangan data batch yang mengekstraksi pola menjadi informasi dari data mentah. Salah satu operasi yang paling umum dalam data mining adalah mengelompokkan data dalam suatu kumpulan data ke dalam kategori atau kelompok tertentu. Teknik *clustering* mengelompokkan objek berdasarkan kesamaan subjektifnya. [5]

Metode clustering K-means merupakan metode pengelompokan yang banyak digunakan. Algoritma ini adalah alat pengelompokan yang paling umum digunakan dalam aplikasi ilmiah dan industri. Merupakan teknik analisis cluster yang bertujuan untuk membagi observasi menjadi k cluster, dimana setiap observasi termasuk dalam cluster yang mempunyai mean terdekat. [6].

Algoritma adalah seperangkat aturan, atau instruksi, yang logis dan sistematis untuk memecahkan masalah atau mencapai tujuan tertentu. Algoritma berfungsi sebagai panduan atau prosedur langkah demi langkah yang harus diikuti untuk mendapatkan hasil yang diinginkan. Algoritma dapat diterapkan di banyak bidang, termasuk matematika, ilmu komputer, dan bidang pengetahuan lainnya. [7]

K-Means adalah metode clustering yang dibutuhkan: sedikit usaha. Pertama, nomor *cluster* c juga menentukan hub mana yang diterima. Objek acak apa pun dengan pusat centroid awal dapat dipilih, atau k objek pertama berturut-turut juga dapat berfungsi sebagai pusat peluncuran. Namun, jika terdapat karakteristik fungsional atau variabel penting yang sangat mempengaruhi pengelompokan hasil dalam kasus ini, maka normalisasi data akan menjadi tugas prapemrosesan yang penting untuk meningkatkan atau mengontrol variabilitas data. Langkah-Langkah-langkah algoritma K-means:

1. Tentukan jumlah *cluster* (k) dan tentukan pusat *cluster* secara acak.
2. Hitung jarak seluruh data ke pusat *cluster*.
3. Mengelompokkan data ke dalam cluster-cluster yang mempunyai jarak terpendek
4. Hitung pusat *cluster* baru.
5. Ulangi langkah 2 (dua) hingga 4 (empat) untuk menghindari pemindahan data ke *cluster*

lain.

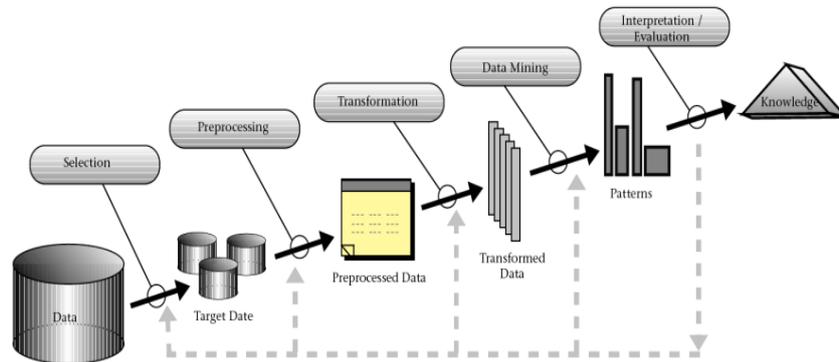
Pajak daerah menurut Undang-Undang Nomor 28 Tahun 2009 adalah wajib kepada daerah yang dibayarkan secara hukum oleh individu atau badan hukum, tidak mendapat imbalan langsung, dan digunakan sebagai keperluan daerah untuk rakyat. Pajak daerah berasal dari pajak daerah itu sendiri atau dari pajak provinsi yang akan dikembalikan kepada daerah untuk memenuhi kebutuhan daerah. [8]

Davies Bouldin Index (DBI) adalah proses yang dipakai untuk mengevaluasi cluster yang dikenalkan oleh David L. Davies dan Donald W. Bouldin. DBI adalah suatu proses yang dipakai untuk mengevaluasi validitas kelompok dengan metode kelompokan. Pengukuran DBI dapat mengoptimalkan jarak antara cluster C_i dan C_j dan pada waktu yang sama meminimalkan jarak antara titik-titik di kelompok.[9]

2. METODE PENELITIAN

Teknik pengumpulan data dalam penelitian ini adalah observasi, subjek penelitian ini adalah Pemerintah Kota Medan Selatan. Data primer adalah jenis data yang digunakan. Data primer berarti data yang dikumpulkan langsung dari Kantor UPT Samsat Medan Selatan menggunakan tinjauan literatur. Tinjauan literatur penelitian ini adalah jurnal terakreditasi dan data pendukung.

Dalam Analisis dan diskusikan isu-isu yang muncul, karena itu penelitian ini menggunakan data kuantitatif. Proses pengambilan data yang akan digunakan untuk data mining dilakukan secara bertahap. Setelah mendapatkan data mentah kemudian dilakukan seleksi dan diolah menjadi sebuah informasi. Pada tahap ini, peneliti menggunakan tahapan Knowledge Discovery In Databases (KDD). KDD adalah proses menemukan informasi yang bermakna dari data yang dikumpulkan dalam database besar. Langkah-langkah KDD yaitu sebagai berikut.[10]



Gambar 1. Tahapan KDD

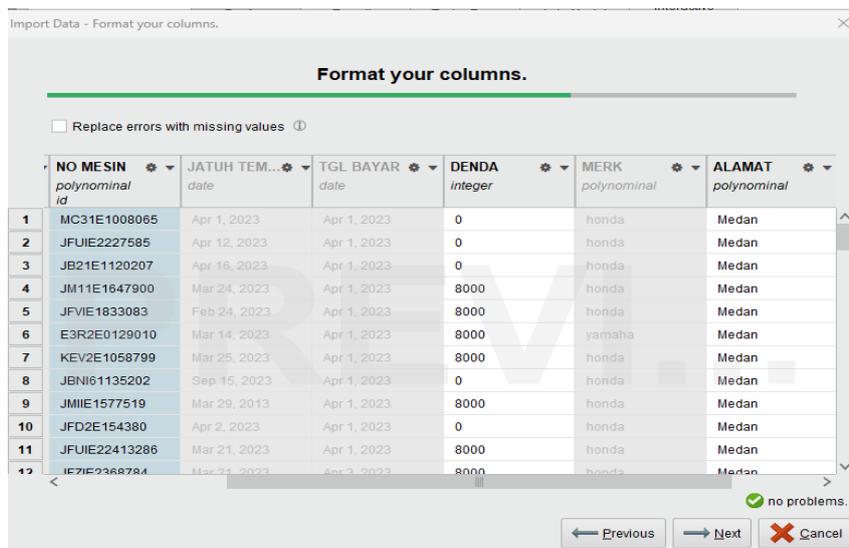
Adapun penjelasan tahapan penelitian yaitu :

1. Data Selection (pemilihan data), diseleksi dengan memilih atribut apa yang diinginkan.
2. Data Cleaning, data-data yang ada akan mengalami pembersihan terhadap data yang tidak diperlukan atau mungkin bernilai null.
3. Transformasi Data (transformasi data), data yang bisa diubah menjadi bentuk yang sesuai untuk di-mining
4. Data Mining, tahap utama dari metode pengumpulan data adalah data mining, yang merupakan identifikasi persyaratan umum yang disepakati untuk mengambil metode tertentu.
5. Interpretasi model Informasi yang diperoleh dari metode data mining yang disediakan dalam bentuk yang cepat dipahami oleh pihak yang memerlukan.

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Pada percobaan awal, sumber data yang telah diperoleh lalu dibedakan dimana data terpakai dan data tidak terpakai. Data yang diperlukan yaitu data yang relevan variabel perhitungan digunakan untuk dianalisis.



Gambar 2. Pemilihan Data

Penelitian ini menggunakan atribut no mesin, denda dan keterangan tepat waktu. Sedangkan atribut alamat digunakan sebagai tambahan untuk menunjang pemodelan data nanti.

3.2 Data Cleaning

Proses menghapus data, memeriksa ketidakcocokan, dan memperbaiki kesalahan adalah bagian dari proses pembersihan, yang bertujuan untuk menurunkan keakuratan hasil data mining yang akan datang.

Row No.	NO MESIN	DENDA	ALAMAT	KETERANGAN
1	MC31E10080...	0	Medan	Tepat waktu
2	JFUIE2227585	0	Medan	Tepat waktu
3	JB21E11202...	0	Medan	Tepat waktu
4	JM11E16479...	8000	Medan	Tidak tepat w...
5	JFVIE1833083	8000	Medan	Tidak tepat w...
6	E3R2E01290...	8000	Medan	Tidak tepat w...
7	KEV2E10587...	8000	Medan	Tidak tepat w...
8	JBNi61135202	0	Medan	Tepat waktu
9	JMIIE1577519	8000	Medan	Tidak tepat w...
10	JFD2E154380	0	Medan	Tepat waktu
11	JFUIE224132...	8000	Medan	Tidak tepat w...
12	JFZIE2368784	8000	Medan	Tidak tepat w...
13	JFZIE2327040	0	Medan	Tepat waktu
14	4D7746975	8000	Medan	Tidak tepat w...
15	JF22IE11697...	0	Medan	Tepat waktu

Gambar 3. Hasil Data *Cleansing*

Pada tahap ini menggunakan operator Replace missing values pada saat proses memasukkan data kedalam rapidminer. Dan hasil dari data cleaning yaitu tidak ada data yang kosong dan missing, sehingga data wajib pajak tetap sebanyak 146 data.

3.3 Data Transformation

Proses ini merupakan jenis informasi yang dicari dalam database menentukan seberapa kreatif proses ini. Selanjutnya, data yang telah dipilih dan dibersihkan akan diubah menjadi bentuk yang sesuai dengan algoritma yang dipakai. Pada metode ini menggunakan operator karena data yang disatukan masih bervariasi nominal, diubah dari nominal ke numerik.

Row No.	NO MESIN	ALAMAT - M.	KETERANGA...	KETERANGA...	KETERANGA...	KETERANGA...	DENDA
1	MC31E10080...	1	1	0	0	0	0
2	JFUIE2227585	1	1	0	0	0	0
3	JB21E11202...	1	1	0	0	0	0
4	JM11E16479...	1	0	1	0	0	8000
5	JFVIE1833083	1	0	1	0	0	8000
6	E3R2E01290...	1	0	1	0	0	8000
7	KEV2E10587...	1	0	1	0	0	8000
8	JBNi61135202	1	1	0	0	0	0
9	JMIIE1577519	1	0	1	0	0	8000
10	JFD2E154380	1	1	0	0	0	0
11	JFUIE224132...	1	0	1	0	0	8000
12	JFZIE2368784	1	0	1	0	0	8000
13	JFZIE2327040	1	1	0	0	0	0
14	4D7746975	1	0	1	0	0	8000
15	JF22IE11697...	1	1	0	0	0	0

Gambar 4. Hasil *Transformasi* Data

3.4 Data Mining

Proses penambangan data adalah proses menemukan pola atau informasi menarik pada sejumlah data tertentu. Ada banyak teknik, metode, atau algoritma penambangan data yang berbeda.

Row No.	NO MESIN	cluster	ALAMAT = M...	KETERANGA...	KETERANGA...	KETERANGA...	KETERANGA...	DENDA
1	MC31E10080...	cluster_3	1	1	0	0	0	0
2	JFUIE2227585	cluster_3	1	1	0	0	0	0
3	JB21E11202...	cluster_3	1	1	0	0	0	0
4	JM11E16479...	cluster_1	1	0	1	0	0	8000
5	JFVIE1833083	cluster_1	1	0	1	0	0	8000
6	E3R2E01290...	cluster_1	1	0	1	0	0	8000
7	KEV2E10587...	cluster_1	1	0	1	0	0	8000
8	JBNI61135202	cluster_3	1	1	0	0	0	0
9	JMIE1577519	cluster_1	1	0	1	0	0	8000
10	JFD2E154380	cluster_3	1	1	0	0	0	0
11	JFUIE224132...	cluster_1	1	0	1	0	0	8000
12	JFZIE2368784	cluster_1	1	0	1	0	0	8000
13	JFZIE2327040	cluster_3	1	1	0	0	0	0
14	4D7746975	cluster_1	1	0	1	0	0	8000

Gambar 5. Tampilan Data Mining

Sesuai dengan metode clustering dengan menerapkan algoritma K-Means dimana nilai $k=4$, max runs sebanyak 10 kali putaran dengan jenis measure types nya adalah NumericalMeasure dengan menggunakan metode perhitungan EuclideanDistance dan max optimization steps nya berjumlah 100.

3.5 Interpretation/Evaluation

Berdasarkan hasil data mining algoritma K-Means didapat Performance sebagai berikut.

```
PerformanceVector:  
Avg. within centroid distance: -22535214.861  
Avg. within centroid distance_cluster_0: -184960000.320  
Avg. within centroid distance_cluster_1: -11408163.334  
Avg. within centroid distance_cluster_2: -42193906.006  
Avg. within centroid distance_cluster_3: -0.094  
Davies Bouldin: -0.294
```

Gambar 6. Hasil Performance Vektor

Lalu proses berikutnya kita memakai operator Cluster Distance Performance yang memungkinkan kita melihat Davies Bouldin Index (DBI) dan nilai rata-rata cluster jarak pusat berdasarkan jarak terdekat yang dikelompokkan dengan algoritma K-Means.

Tabel 1. Percobaan menentukan nilai k

Nilai k	AVG	DBI
2	123818348.38	0.297
3	41965593.07	0.304
4	22535214.861	0.294

Selanjutnya mencoba 4 kali percobaan untuk menemukan nilai k terbaik, percobaan ketiga menghasilkan nilai k = 4 ternyata lebih optimal dibandingkan percobaan pertama dan kedua karena nilai DBI mendekati 0 yaitu 0,294.



Gambar 7. Hasil Cluster yang sudah membayar pajak

Pada Gambar 8, cluster merupakan hasil setiap blok pembayaran pajak kendaraan bermotor. Keanggotaan cluster pada kemiripan, pemangatan hasil clustering dikelompokkan berdasarkan tingkat kepatuhan mereka. Hasil clustering menunjukkan bahwa wajib pajak di blok cluster 0 yang berwarna jingga memiliki tingkat kepatuhan pembayaran pajak yang sangat rendah, dengan nilai 10 wajib pajak yang tidak tepat waktu dengan denda yang paling tinggi yaitu antara 60-100k. Tingkat kesadaran wajib pajak pada blok cluster 1 yang berwarna hijau bisa dikatakan sedang, yang artinya tidak tepat waktu tetapi denda nya masih rendah yaitu antara 8-20k dengan memiliki 56 wajib pajak. Pada blok yang berwarna hitam cluster 2 memiliki tingkat kepatuhan pajak yang rendah dengan denda yang cukup tinggi yaitu antara 24-56k memiliki 19 wajib pajak. Cluster terakhir yang dihasilkan adalah blok cluster 3 dengan tingkat kepatuhan tinggi dan tidak memiliki denda sebanyak 61 wajib pajak. Berikut tabel anggota cluster berdasarkan tingkat kepatuhannya.

Tabel 2. Klaster Tingkat Kepatuhannya

Cluster	Anggota	Tingkat kepatuhan
Cluster 0	10	Sangat rendah
Cluster 1	56	Sedang

Cluster 2	19	Rendah
Cluster 3	61	Tinggi

4. KESIMPULAN DAN SARAN

Penelitian ini menggunakan data dari 146 wajib pajak yang membayar pajak untuk pembayaran pajak kendaraan bermotor. Studi ini mengidentifikasi empat kelompok kepatuhan pajak kendaraan bermotor: sangat rendah, rendah, sedang, dan tinggi. cluster 0 dinilai memiliki tingkat kepatuhan yang sangat rendah, cluster 2 memiliki tingkat kepatuhan yang rendah, cluster 1 memiliki tingkat kepatuhan yang rata-rata, dan cluster 3 memiliki tingkat kepatuhan yang tinggi. Hasil penelitian ini bertujuan untuk memberikan rekomendasi atau model yang dapat digunakan untuk menggunakan algoritma K-Means pada data pembayaran pajak kendaraan bermotor dalam upaya menentukan tingkat kepatuhan wajib pajak. untuk pemerintah dapat mengetahui cara untuk meningkatkan kesadaran para wajib pajak yang sangat rendah dan tidak patuh.

DAFTAR PUSTAKA

- [1] Siahaan, Marihot P. (2013). Pajak Daerah dan Retribusi Daerah, Jakarta : Raja Grafindo
- [2] Kusumawati, I. N., & Rachman, A. N. (2021). Analisis Pengaruh Wajib Pajak dalam Membayar Pajak Kendaraan Bermotor. *Jurnal Ekonomi-QU*, 11(1), 1–20.
- [3] I. Irwan, S. Sidjara, and A. P. Aryati. (2022). Pengelompokan Jenis Penerimaan Pajak di Kota Makassar Menggunakan Fuzzy Clustering. *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 10, no. 1. 98–102
- [4] Sari, Desti Puspita, Shofa Shofia Hilabi, and Agustia Hananto. (2023). Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama. *SMARTICS Journal* 9.1: 14-19.
- [5] Nasution, M. Z., & Hasibuan, M. S. (2020). Pendekatan Initial Centroid Search Untuk Meningkatkan Efisiensi Iterasi Klustering K-Means. *Techno. Com*, 19(4), 341-352.
- [6] Syahputra, YH, & Hutagalung, J. (2022). Kelas unggulan untuk meningkatkan prestasi belajar siswa menggunakan algoritma K-means. *Sinkron : jurnaldan penelitian teknik informatika*, 7 (3), 891-899
- [7] Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2022). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*.
- [8] Mukmin, M. N., & Maemunah, S. (2019). Pengelolaan Dana Pemerintah Desa: Kajian Pada Kecamatan Babakan Madang, Sukaraja Dan Ciawi. *Jurnal Akunida*, 4(2), 73-85.

-
- [9] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo. (2019). Perbandingan Algoritma K-Means dan K-Medoids dalam Klasterisasi Produk Asuransi Perusahaan Nasional. *Jurnal Ilmiah FIFO*. vol. 11, no. 2, p. 208
- [10] Haris Kurniawan, Sarjon Defit, and Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2.