
Penerapan Algoritma K-Means Untuk Menentukan Kelas Besar Tunggakan Pajak Kendaraan Bermotor Pada SAMSAT Parigi Moutong

Sukardi¹, Anwar S. Panyili² and Nur Alinuddin Kaharu³

¹²³STMIK Adhi Guna
kardi0673@gmail.com

Abstract

Tax is a mandatory contribution imposed on citizens and serves as a significant source of state revenue, used for development as mandated by law. One major tax type in Indonesia is the motor vehicle tax, managed by the One-Stop Administrative System (SAMSAT) at the district level. Each district, including Parigi Moutong (Parimo), has a SAMSAT office responsible for managing motor vehicle tax, including arrears. The high number of registered vehicles at SAMSAT Parimo corresponds to the large volume of data managed, including tax arrears data. Motor vehicle tax arrears vary for each taxpayer and influence policy decisions by SAMSAT. To address the variation in tax amounts and required actions, data classification based on tax arrears is necessary. The advancement of technology and computational algorithms allows large data sets to be processed effectively through data mining methods. This research explores the application of data mining techniques to classify motor vehicle tax arrears and assist SAMSAT Parimo in optimizing tax management.

Keywords: *Tax; Motor Vehicle Tax; SAMSAT; Tax Arrears; Data Mining*

Abstrak

Pajak merupakan kontribusi wajib yang menjadi sumber utama pendapatan negara untuk pembangunan. Salah satu pajak utama di Indonesia adalah pajak kendaraan bermotor yang dikelola oleh Sistem Administrasi Manunggal Satu Atap (SAMSAT) di tingkat kabupaten, termasuk Kabupaten Parigi Moutong (Parimo). Pengelolaan pajak kendaraan bermotor mencakup data penunggakan yang beragam pada setiap wajib pajak, memengaruhi kebijakan yang harus diambil. Banyaknya kendaraan terdaftar di SAMSAT Parimo menghasilkan data besar, termasuk penunggakan pajak yang membutuhkan pengelolaan efektif. Variasi jumlah tagihan dan tindakan yang diperlukan mendorong perlunya klasifikasi data berdasarkan tagihan pajak. Dengan perkembangan teknologi dan algoritma komputasi, pengolahan data besar dapat dilakukan melalui metode data mining. Penelitian ini membahas penerapan teknik data mining untuk mengklasifikasikan penunggakan pajak kendaraan bermotor guna mendukung pengelolaan pajak yang optimal di SAMSAT Parimo.

Kata Kunci: **Pajak; Pajak Kendaraan Bermotor; SAMSAT; Penunggakan Pajak; Data Mining**

Pendahuluan

Pajak merupakan pungutan yang wajib dipenuhi oleh warga negara dan menjadi salah satu sumber pendapatan negara yang dimanfaatkan untuk pembangunan sesuai amanah Undang-undang. Pendapatan dari sektor pajak merupakan pendapatan terbesar dari jumlah seluruh pendapatan negara. salah satu jenis pajak yang diberlakukan di Indonesia adalah pajak kendaraan bermotor yang dikelola pada satuan Sistem Administrasi Manunggal Satu Atap (SAMSAT). Pajak kendaraan bermotor adalah pajak yang diberikan kepada pengguna kendaraan bermotor yang yang berbeda besarnya tergantung jenis kendaraan dan dipungut setiap tahun. Pengelolaan pajak kendaraan bermotor oleh SAMSAT dilakukan pada tingkat kabupaten, dengan demikian setiap kabupaten terdapat kantor cabang SAMSAT masing-masing, termasuk Kabupaten Parigi Moutong (Parimo).

Pengelolaan pajak pada kantor samsat mencakup penunggakan. Dengan banyaknya kendaraan yang terdaftar pada samsat parimo, berbanding lurus dengan jumlah data yang dikelola oleh pihak SAMSAT termasuk data penunggakan pajak kendaraan bermotor terdaftar. Penunggakan pajak kendaraan bermotor tentu saja bervariasi jumlah besarnya pada setiap wajib pajak. Besar dan kecilnya tagihan pajak menentukan kebijakan yang perlu diambil oleh pihak SAMSAT. Dengan banyaknya variasi jumlah dan tindakan yang perlu diambil mengikuti besaran tagihan tersebut, maka perlu pemilahan data berdasarkan jumlah tagihan pajak. Dengan berkembangnya teknologi dan algoritma komputasi, pemilahan data yang besar dapat ditangani dengan penerapan metode datamining.

Metode

Dalam penelitian ini jenis penelitian yang digunakan yaitu penelitian terapan. Penelitian terapan atau applied research dilakukan berkenaan dengan kenyataan-kenyataan praktis, penerapan, dan pengembangan ilmu pengetahuan yang dihasilkan oleh penelitian dasar dalam kehidupan nyata. Penelitian terapan berfungsi untuk mencari solusi tentang masalah-masalah tertentu. Tujuan utama penelitian terapan adalah pemecahan masalah sehingga hasil penelitian dapat dimanfaatkan untuk kepentingan manusia baik secara individu atau kelompok maupun untuk keperluan industri atau politik dan bukan untuk wawasan keilmuan semata.

Untuk mendapatkan data-data yang diperlukan dalam penyusunan proposal skripsi ini dilakukan dengan menggunakan dua jenis data, yaitu sebagai berikut: Data primer merupakan data yang dikumpulkan peneliti secara langsung dari tempat objek penelitian dengan cara peneliti melakukan wawancara langsung dengan staff di kantor Samsat Parimo. Data sekunder merupakan sebuah informasi yang sebelumnya ada dan secara sengaja dikumpulkan oleh peneliti dan dipergunakan untuk melengkapi kebutuhan data yang dibutuhkan peneliti. Data ini biasanya diperoleh dari berbagai macam sumber baik dari objek penelitian maupun jurnal-jurnal penelitian terkait, buku-buku, literatur yang relevan dengan permasalahan dari peneliti baik gambaran umum tempat penelitian, serta informasi lain yang terhubung dengan berbagai penelitian yang telah dilakukan.

Hasil dan Pembahasan

Penelitian ini diawali dengan proses pengumpulan data wajib pajak kendaraan bermotor di Kabupaten Parigi Moutong (Parimo). Data yang terkumpul mencakup berbagai informasi penting, seperti identitas pemilik kendaraan, spesifikasi kendaraan,

serta parameter yang relevan dengan perhitungan pajak kendaraan bermotor. Dataset yang diperoleh terdiri dari 19 atribut, antara lain “nama”, “nopol” (nomor polisi), “hp” (nomor telepon), “alamat”, “merk”, “tipe”, “tahun buat”, “warna TNKB” (Tanda Nomor Kendaraan Bermotor), “bobot”, “nilai jual”, “dasar PKB” (Pajak Kendaraan Bermotor), “SKPD” (Surat Ketetapan Pajak Daerah), “lama tunggakan”, “pokok”, “denda”, “total PKB”, dan “status”.

Table 1.
Dataset Dengan Atribut Awal

Nama	Nopol	Denda	Total PKB	Status
Rahmi	DN1201DG	421.313	4.196.273	Ada dan beroperasi
Arifin	DN1704DE	444.938	4.360.388	Ada dan beroperasi
Hanapi	DN2083DO	0	184.440	Ada dan beroperasi
Nurdin	DN5111DP	0	302.940	Ada dan beroperasi
Nurmilasari	DN4978DT	0	254.040	Ada dan beroperasi
Wahida m.hi.thayib	DN3438DW	52.875	501.255	Ada dan beroperasi
Mohammad nur sa,ban	DN4270DP	43.125	429.525	Ada dan beroperasi
M.yunus s.	DN2669DF	31.500	308.700	Ada dan beroperasi
Ardi	DN5508DS	94.500	918.540	Ada dan beroperasi
Marmin.l	DN2250DK	31.875	297.075	Ada dan beroperasi
Kantor desa tampiala	DN2026T	15.625	144.375	Ada dan beroperasi

Pada proses input data, dataset yang sudah direduksi atributnya secara teknis ditampung ke dalam file excel untuk dapat diproses pada program yang dikembangkan. Hasil dari pengurangan reduksi dapat dilihat pada tabel berikut.

Table 2.
Dataset Hasil Reduski Atribut

Thn. Buat	Bobot	Nilai Jual	Dasar PKB	Usia SKPD	Lama Tunggakan	Total PKB
2012	1,05	113.000.000	118.650.000	721	1	4.360.388
2012	1,05	113.000.000	118.650.000	498	0	1.957.725
2012	1,05	111.000.000	116.550.000	821	1	4.143.353
2012	1	13.100.000	13.100.000	388	0	231.870
2012	1	11.100.000	11.100.000	705	1	407.925
2012	1	12.600.000	12.600.000	699	1	463.050
2012	1	11.000.000	11.000.000	818	1	391.050
2012	1,085	62.000.000	67.270.000	390	0	1.190.679
2012	1	11.200.000	11.200.000	764	1	404.880
...
2021	1,085	93.000.000	100.905.000	543	0	1.634.661
2021	1	14.800.000	14.800.000	364	0	261.960
2021	1	18.000.000	18.000.000	565	0	286.200
2021	1	11.900.000	11.900.000	411	0	207.060
2021	1	13.000.000	13.000.000	439	0	222.300
2021	1	12.400.000	12.400.000	442	0	212.040

Thn. Buat	Bobot	Nilai Jual	Dasar PKB	Usia SKPD	Lama Tunggakan	Total PKB
2021	1	14.300.000	14.300.000	411	0	248.820
2021	1,05	134.000.000	140.700.000	451	0	2.405.970
2021	1,05	117.000.000	122.850.000	504	0	2.027.025
2021	1	12.400.000	12.400.000	521	0	200.880

Proses pembagian kelas menggunakan k-means dilakukan dengan menghitung jarak satu data dengan centroid. Centroid merupakan titik patokan dimana data-data yang dekat dengan centroid digolongkan dalam satu kelas yang sama. Proses dimulai dengan menentukan jumlah K dan posisi centroid setiap kelas. Untuk menentukan jarak satu data dengan data centroid digunakan persamaan Euclidean distance. Berikut dicontohkan proses cluster menggunakan k-means pada 10 data pertama dataset yang kami gunakan.

Table 3
Contoh Dataset

n ₁ :	1,05	113.000.000	118.650.000	721	1	4.360.388
n ₂ :	1,05	113.000.000	118.650.000	498	0	1.957.725
n ₃ :	1,05	111.000.000	116.550.000	821	1	4.143.353
n ₄ :	1	13.100.000	13.100.000	388	0	231.870
n ₅ :	1	11.100.000	11.100.000	705	1	407.925
n ₆ :	1	12.600.000	12.600.000	699	1	463.050
n ₇ :	1	11.000.000	11.000.000	818	1	391.050
n ₈ :	1,085	62.000.000	67.270.000	390	0	1.190.679
n ₉ :	1	11.200.000	11.200.000	764	1	404.880
n ₁₀ :	1	12.600.000	12.600.000	430	0	215.460

K-Means clustering dimulai dengan menentukan jumlah kelas yang akan dihasilkan dan centroid yang digunakan. Pada contoh perhitungan ini, jumlah kelas yang ingin dihasilkan berjumlah 2 dan centroid yang digunakan berasal dari data pertama dan data terakhir.

Menentukan jumlah kelas dan centroid

Jumlah cluster :

K = 2

Centroid1 = [1,05; 113.000.000; 118.000.000; 721; 1; 4.360.388]

Centroid2 = [1; 12.600.000; 12.600.000; 430; 0; 215.460]

Menghitung jarak seluruh dataset ke masing-masing centroid. Perhitungan distance dari setiap data dari centroid1

$$D_{11} = \sqrt{(1,05 - 1,05)^2 + (113000000 - 113000000)^2 + \dots + 0}$$

$$D_{12} = \sqrt{(1,05 - 1,05)^2 + (113000000 - 113000000)^2 + \dots +}$$

$$D_{13} = \sqrt{(1,05 - 1,05)^2 + (111000000 - 113000000)^2 + \dots +}$$

$$D_{14} = \sqrt{(1 - 1.05)^2 + (13100000 - 113000000)^2 + \dots}$$

$$D_{15} = \sqrt{(1 - 1.05)^2 + (11100000 - 113000000)^2 + \dots}$$

$$D_{16} = \sqrt{(1 - 1.05)^2 + (12600000 - 113000000)^2 + \dots}$$

$$D_{17} = \sqrt{(1 - 1.05)^2 + (11000000 - 113000000)^2 + \dots}$$

$$D_{18} = \sqrt{(1.085 - 1.05)^2 + (62000000 - 113000000)^2 + \dots}$$

$$D_{19} = \sqrt{(1 - 1.05)^2 + (11200000 - 113000000)^2 + \dots}$$

$$D_{110} = \sqrt{(1 - 1.05)^2 + (12600000 - 113000000)^2 + \dots}$$

Perhitungan *distance* dari setiap data dari *centroid*₂

$$D_{21} = \sqrt{(1.05 - 1)^2 + (113000000 - 12600000)^2 + \dots} = 146095663,6$$

$$D_{22} = \sqrt{(1.05 - 1)^2 + (113000000 - 12600000)^2 + \dots}$$

$$D_{23} = \sqrt{(1.05 - 1)^2 + (111000000 - 12600000)^2 + \dots}$$

143190749,9

$$D_{24} = \sqrt{(1 - 1)^2 + (13100000 - 12600000)^2 + \dots}$$

$$D_{25} = \sqrt{(1 - 1)^2 + (11100000 - 12600000)^2 + \dots}$$

$$D_{26} = \sqrt{(1 - 1)^2 + (12600000 - 12600000)^2 + \dots}$$

$$D_{27} = \sqrt{(1 - 1)^2 + (11000000 - 12600000)^2 + \dots}$$

$$D_{28} = \sqrt{(1.085 - 1)^2 + (62000000 - 12600000)^2 + \dots}$$

$$D_{29} = \sqrt{(1 - 1)^2 + (11200000 - 12600000)^2 + \dots}$$

$$D_{210} = \sqrt{(1 - 1)^2 + (12600000 - 12600000)^2 + \dots}$$

Proses yang selanjutnya dilakukan adalah membandingkan nilai dua distance pada data yang sama untuk menentukan kelas dari data tersebut. Sebagai contoh, n₁ memiliki nilai distance dari centroid₁ sebesar 0 dan distance dari centroid₂ senilai 146.095.663,6. Dengan demikian, kelas dari n₁ adalah kelas 1 karna distance terhadap centroid₁ lebih kecil dari distance terhadap centroid₂.

Table 4.
Contoh Hasil Cluster 10 Data Pertama Iterasi 1

n	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	D ₁	D ₂	C
n ₁	1,05	113000000	118650000	721	1	4360388	0	146095663,6	Kls 1
n ₂	1,05	113000000	118650000	498	0	1957725	2402663,01	146047245,7	Kls 1
n ₃	1,05	111000000	116550000	821	1	4143353	2908110,074	143190749,9	Kls 1
n ₄	1	13100000	13100000	388	0	231870	145388641,8	707297,1722	Kls 2
n ₅	1	11100000	11100000	705	1	407925	148210102,4	2130033,533	Kls 2
n ₆	1	12600000	12600000	699	1	463050	146088848,8	247590,1461	Kls 2
n ₇	1	11000000	11000000	818	1	391050	148351872,7	2269544,447	Kls 2
n ₈	1085	62000000	67270000	390	0	1190679	72463449,11	73689347,62	Kls 1
n ₉	1	11200000	11200000	764	1	404880	148068864,2	1988939,428	Kls 2
n ₁₀	1	12600000	12600000	430	0	215460	146095663,6	0	Kls 2

Tabel sebelumnya menunjukkan hasil cluster dari 10 data pertama dengan menentukan data pertama sebagai centroid₁ dan data ke-10 sebagai centroid₂. Cluster yang dihasilkan pada tabel tersebut bukan merupakan hasil final dari proses k-means clustering. Proses selanjutnya adalah mengubah nilai centroid kemudian diuji kembali pada menggunakan Euclidean distance. Proses perubahan nilai centroid ini dilakukan terus menerus hingga kelas yang dihasilkan tidak mengalami perubahan. Karna hasil cluster pada tabel sebelumnya merupakan perhitungan jarak pertama dari seluruh proses k-means clustering, maka proses sebelumnya disebut dengan iterasi ke-1.

Nilai centroid baru ditentukan dengan menjumlahkan seluruh nilai atribut yang sejenis dalam satu kelas dan dibagi dengan jumlah data dalam satu kelas. Lebih mudah menghasilkan centroid baru dengan memisahkan data dari tiap kelas. Pemisahan data dari tiap kelas dapat dilihat pada 2 tabel berikut :

Table 5.
Data Kelas 1 Iterasi-1

n	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
n ₁	1,05	113000000	118650000	721	1	4360388
n ₂	1,05	113000000	118650000	498	0	1957725
n ₃	1,05	111000000	116550000	821	1	4143353
n ₈	1085	62000000	67270000	390	0	1190679
C ₁	272,0375	99750000	105280000	607,5	0,5	2913036,25

Tabel sebelumnya menunjukkan data pada kelas 1 iterasi pertama dengan nilai centroid baru dapat dilihat pada akhir tabel.

Table 6.
Data Kelas 2 iterasi-1

n	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
n ₄	1	13100000	13100000	388	0	231870
n ₅	1	11100000	11100000	705	1	407925
n ₆	1	12600000	12600000	699	1	463050
n ₇	1	11000000	11000000	818	1	391050
n ₉	1	11200000	11200000	764	1	404880
n ₁₀	1	12600000	12600000	430	0	215460

Tabel sebelumnya menunjukkan data pada kelas 2 iterasi pertama dengan nilai centroid bari dapat dilihat pada akhir tabel. Hasil evaluasi dapat dilihat pada tabel berikut.

Table 7
Evaluasi Performa K-means

Jumlah kelas	Silhouette Score	Davies Bouldin score
K = 2	0.86061	0,35156
K = 3	0.88491	0,3838
K = 4	0.86313	0,4383
Rata-rata	0,86955	0,39122

Kesimpulan

Penilaian performa dilakukan untuk mengetahui seberapa dekat suatu data dengan data yang lain dalam kelas yang sama dan seberapa jauh satu data dengan data yang lain pada kelas yang berbeda. Eksperimen yang dilakukan pada bab sebelumnya menghasilkan dua jenis score yaitu silhouette index dan davies bouldin index. Kedua metode penilaian ini umum digunakan untuk menilai performa k-means clustering. Silhouette Index memiliki rentang nilai dari -1 hingga 1. Semakin mendekati nilai 1 maka semakin baik performa cluster. Pada eksperimen yang telah dilakukan rata-rata silhouette index sebesar 0,86955 dengan nilai silhouette index terbesar dengan nilai 0,88891 dengan K=3.

Daftar Pustaka

- Ardianti, D. (2014). Pelaksanaan pemungutan pajak kendaraan bermotor oleh dinas pendapatan daerah provinsi Jawa Timur (studi kasus di kantor dinas pendapatan daerah provinsi Jawa Timur unit pelaksana teknis dinas (UPTD) Malang Kota). *Kumpulan Jurnal Mahasiswa Fakultas Hukum*.
- Dayat Suryana. (n.d.). Controls Visual Basic Jilid 1. Pengertian dataset adalah representasi data yang disimpan di memori dalam kondisi tidak terhubung (disconnected).
- Gustian, D., Rustaman, K., Yunita, A. N., Neng, G. N., Noviansyah, R., & Junfithrana, A. P. (2019). Pemantauan pola pengunjung perpustakaan dengan metode FP-Growth guna meningkatkan fungsi perusahaan. *Jurnal Informatics for Educators and Professionals*.
- Fauzy, M., et al. (2016). Penerapan metode association rule menggunakan algoritma Apriori pada simulasi prediksi hujan wilayah Kota Bandung. *Jurnal Ilmiah Teknologi Informasi Terapan*.
- Prasetyo, E. (2012). *Data Mining: Konsep dan aplikasi menggunakan MATLAB*. Yogyakarta: Andi.

- Siregar, Y. A., Saryadi, S., & Listyorini, S. (2012). Pengaruh pelayanan fiskus dan pengetahuan perpajakan terhadap kepatuhan wajib pajak (studi empiris terhadap wajib pajak di Semarang Tengah). *Jurnal Ilmu Administrasi Bisnis*, 1(2), 295–304.
- Suprihatin. (2011). Klastering k-means untuk penentuan nilai ujian. *JUSI*, 1(1).