

BAB I

PENDAHULUAN

1.1. Latar Belakang

Penerapan teknologi informasi data mining di Kantor Desa, khususnya dalam layanan administrasi kantor desa, menjadi sangat penting untuk memastikan efisiensi dan keakuratan layanan informasi (Astuti, 2022). Seperti halnya, dalam pelayanan Pajak Bumi dan Bangunan (PBB) Perdesaan dan Perkotaan yang merupakan pajak atas tanah dan bangunan yang dimiliki, dikuasai, dan/atau dimanfaatkan oleh individu atau organisasi, kecuali kawasan yang digunakan untuk kegiatan perkebunan, perhutanan, dan pertambangan, menurut Undang-Undang nomor 28 tahun 2009 tentang Pajak Daerah dan Retribusi Daerah . Dalam pasal 28 ini, "bumi" berarti permukaan bumi yang terdiri dari tanah, perairan pedalaman, dan laut wilayah kabupaten atau kota. Sedangkan "bangunan" berarti konstruksi teknik yang ditanam atau diletakkan secara permanen pada tanah, perairan pedalaman, atau laut. Pajak mempunyai dua fungsi: pengendalian anggaran dan pengawasan. Pajak berfungsi sebagai alat pemerintah untuk memperoleh dana dari masyarakat yang dapat digunakan untuk berbagai kepentingan pemerintah. Fungsi anggaran ini menuntut warga negara untuk menyadari kewajibannya sebagai wajib pajak dan memenuhinya dengan disiplin. Hal ini menentukan seberapa baik seorang wajib pajak mematuhi peraturan di bidang tertentu. Banyaknya data yang disimpan sebanding dengan jumlah wajib pajak di daerah tersebut. (Putri dkk., 2022).

Di desa pandanarum terdapat jenis pajak bumi dan bangunan yang datanya terdapat pada buku daftar himpunan ketetapan pajak (DHKP) yang berjumlah 4576

jiwa penduduk wajib pajak yang terbagi dari dua jenis pajak bangunan dan sawah. Buku DHKP berisi pencatatan setiap rumah dan pemilik tanah untuk mengatur wajib pajak yang berikan pada setiap pemilik tanah. Pendataan buku ini dilakukan secara setiap tahun. selain itu, Adapun Buku tanah yakni *Letter C* merupakan dari berbagai bentuk kepemilikan tanah yang ada di Indonesia. Ini dikeluarkan oleh Badan Pertanahan Nasional dan biasanya diterima dari desa Buku *Letter C* Desa menunjuk pemilik tanah yang sebenarnya, atau subjeknya. Hal ini antara lain mencakup uraian tentang sejarah properti, letak properti, kavling, luas properti, batas properti dan nomor pemilik properti, serta nomor pajak properti. Kelurahan pandanarum sendiri memiliki Luas sebesar 369 hektar dan memiliki penduduk sebanyak 8900 jiwa dengan 2561 kepala keluarga (Santi dkk., 2023).

Menurut hasil observasi yang dilaksanakan Desember 2023 Buku *Letter C* dan Daftar Himpunan Ketetapan Pajak (DHKP) tersebut masih dilakukan pencatatan secara manual sehingga membuat data wilayah dusun wajib pajak tidak teratur dan tidak urut, ini menimbulkan kendala dalam pelayanan petugas desa untuk mencari beberapa wilayah dusun yang wajib pajak dengan pajak terhutang. serta, membutuhkan waktu yang relatif lama. Sehingga diperlukannya metode pengelompokkan *clustering*.

Berdasarkan permasalahan diatas diperlukannya pengelompokkan wilayah dusun wajib pajak berdasarkan pajak terhutang. Pengelompokkan merupakan salah satu metode data mining untuk mengelompokkan data dengan karakteristik yang sama ke suatu kelompok dan data dengan karakteristik yang berbeda ke kelompok yang lain. Tujuan pengelompokkan cluster ini adalah untuk menemukan pola, titik,

objek, dan pengelompokan kumpulan dokumen. Objek dalam kelompok cluster yang sama mempunyai kesamaan antar kelompok dan berbeda dengan objek dalam kelompok cluster lainnya (Wahyudi & Pujiastuti, 2022). Metode *Clustering* yang digunakan dalam penelitian ini yaitu menggunakan Algoritma *K-Means* dan *K-Medoids*. Algoritma pengelompokan *K-Means* adalah cenderung lebih cepat dan lebih efisien dari segi komputasi karena perhitungannya menggunakan perhitungan rata-rata untuk menentukan centroid. *K-Medoids* adalah algoritma clustering yang bertujuan untuk meminimalkan kriteria kesalahan *absolut* dan lebih tahan terhadap outlier karena memilih medoid sebagai pusat klaster, yang merupakan anggota aktual dari dataset. Ini menjadikannya lebih representatif, khususnya jika "Pajak Terhutang" atau "Usia" memiliki distribusi yang lebar atau *skewed*.

Adapun tujuan melakukan perbandingan yaitu menentukan algoritma terbaik dalam *clustering* wilayah dusun dengan menganalisis pola pengelompokan data untuk pihak kantor desa agar bisa meningkatkan dan mengelompokkan wilayah dusun dan tujuannya mempermudah pekerjaan petugas kantor desa pandanarum dalam *clustering* wilayah dusun wajib pajak dengan indikator pajak terhutang. Algoritma *K-Means* dan *K-Medoids* ini dikenal karena sangat akurat dan efisien ketika berhadapan dengan objek yang besar. Mardalius menyatakan bahwa kelebihan fleksibilitas metode *K-Means* dan *K-Medoids* adalah karena pengguna dapat memilih berapa banyak cluster yang akan dihasilkan (Prasetyaningrum & Susanti, 2023). Penelitian ini menggunakan Metode pengujian *Silhouette Coefficient* untuk menentukan kualitas dan kekuatan *cluster* (Zahrotun dkk., 2023) dan menggunakan Metode pengujian *Elbow* Hasil dari pengelompokan wilayah dusun

wajib pajak akan dibagi menjadi tiga kluster pajak terhutang: kluster wilayah tertinggi, kluster wilayah sedang, dan kluster wilayah terendah.

Terdapat beberapa penelitian sebelumnya yang serupa dilakukan oleh (Hidayat & Fitriana, 2022) Implementasi *K-Means* Dan *K-Medoids* Dalam Pengelompokan Wilayah Potensial Produksi Daging Ayam. Selanjutnya penelitian oleh (Prasetyaningrum & Susanti, 2023) Perbandingan Algoritma *K-Means* Dan *K-Medoids* Untuk Pemetaan Hasil Produksi Buah-Buahan.

Berdasarkan latar belakang masalah dan pemaparan studi literatur, terdapat beberapa penerapan Algoritma metode Clustering yang digunakan yaitu Algoritma *K-Means* dan Algoritma *K-Medoids*. Dengan dasar ini, maka peneliti memutuskan untuk mengambil judul "Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk *Clustering* Wilayah Dusun Berdasarkan Pajak Terhutang pada Dokumen Ketetapan Pajak".

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil pengelompokkan berdasarkan indikator pajak terhutang pada Kecamatan Sutojayan menggunakan Algoritma *K-Means* dan *K-Medoids* ?
2. Bagaimana hasil perbandingan antara Algoritma *K-Means* dan *K-Medoids* Dalam pengelompokkan Dusun di Kecamatan Sutojayan berdasarkan indikator pajak terhutang pada tahun 2023 ?

3. Bagaimana hasil Evaluasi Pengujian Algoritma *K-Means* dan *K-Medoids* dalam pengelompokan pajak terhutang dengan menggunakan metode *Sum Squared Error (SSE)* dan *Silhouette Coefficient*?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui hasil pengelompokan Wilayah Dusun di kecamatan sutojayan berdasarkan indikator pajak terhutang menggunakan Algoritma *K-means* dan *K-medoids*.
2. Untuk mengetahui hasil perbandingan kinerja antara Algoritma *K-means* dan *K-medoids* dalam pengelompokan Wilayah Dusun di kecamatan sutojayan berdasarkan indikator pajak terhutang pada tahun 2023.
3. Untuk mengetahui hasil evaluasi kinerja perbandingan Algoritma *K-means* dan *K-medoids* menggunakan pengujian *Sum Squared Error (SSE)* dan *Silhouette Coefficient*.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi semua pihak yang terlibat, baik lembaga, masyarakat, maupun penulis sendiri. Adapun manfaat yang diharapkan yaitu:

1. Diharapkan dapat menjadi menambah wawasan dan pengetahuan mengenai kinerja pengelompokan indikator wilayah dusun pajak terhutang menggunakan Algoritma *K-means* dan *K-medoids Clustering*

2. Mengevaluasi Efektivitas Algoritma *K-means* dan *K-medoids Clustering* dalam mengklasterkan Wilayah Dusun berdasarkan pajak terhutang
3. Diharapkan penelitian ini dijadikan bahan referensi Universitas agar penelitian ini dapat dijadikan referensi untuk penelitian selanjutnya.
4. Diharapkan penelitian ini dapat membantu pihak petugas Instansi Kantor Desa Pandanarum dalam mengelompokkan wilayah dusun berdasarkan pajak terhutang.

1.5. Batasan Masalah

Batasan Masalah pada penelitian ini mencakup 3 aspek yaitu metode, data, dan *tools* yaitu sebagai berikut:

1. Metode

Adapun metode yang digunakan dalam penelitian ini yaitu:

- a) Pengolahan pengelompokan Algoritma *K-Means* dan *K-Medoids*.
- b) Membandingkan kinerja hasil analisa Algoritma *K-Means* dan *K-Medoids* dengan melakukan simulasi beberapa skenario pada algoritma *k-means* dan *k-medoids*
- c) Penentuan Cluster terbaik menggunakan Metode Elbow
- d) Menguji Algoritma *K-Means* dan *K-Medoids* menggunakan metode evaluasi *Sum Squared Error* dan *Silhouette Index*

2. Data

Adapun sumber data yang saya gunakan untuk proses Perbandingan Algoritma sebagai berikut:

- a) Data yang digunakan yaitu data sekunder yaitu data sekunder yaitu data Daftar Himpunan Ketetapan pajak (DHKP) tahun 2023 dan Data Kependudukan
- b) Variable yang digunakan dalam penelitian ini yaitu ada yakni objek pajak, pajak terhutang, usia, pendidikan terakhir
- c) Studi Kasus: Kantor Desa Pandanarum Kecamatan Sutojayan
- d) Jumlah *cluster* yang ditentukan ada 3 yakni: Pajak Terhutang Tinggi, Sedang, dan Rendah.

3. *Tools*

Adapun *tools* yang digunakan untuk proses perbandingan Algoritma sebagai berikut:

- a) Menggunakan *Google Colab* IDE untuk melakukan pengelompokkan
- b) Menggunakan bahasa pemrograman *Python*
- c) Menggunakan *Spreadsheet Microsoft Excel* untuk pengolahan datanya

1.6. **Sistematika Penulisan**

Penulisan proposal ini dibagi lima bab yaitu:

BAB I PENDAHULUAN

Dalam bab ini berisi latar belakang yang berisi tentang fenomena permasalahan terkait buku desa *letter C* kecamatan sutojayan, Rumusan masalah, Tujuan Penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Dalam bab ini berisi tentang landasan teori Dasar Pajak bumi dan bangunan, Pajak, Data Mining, Algoritma *Clustering K Means* dan *K Medoids*, normalisasi data, teori slovin, Evaluasi *Sum Squared Error* dan *Silhouette Index*, dan Tools yang digunakan. Landasan Teori ini bersumber dari buku, Jurnal, Artikel. Selain itu, pada bab ini berisikan Kajian Literatur pendukung penelitian ini

BAB III METODOLOGI PENELITIAN

Bab ini berisi waktu dan Lokasi Penelitian, Jenis Penelittian yang digunakan dimana metode yang digunakan yaitu deskriptif kuantitatif, Pengumpulan Data, dan Tahap-Tahap Penelitian, yang digunakan dalam melakukan simulasi.

BAB IV HASIL PEMBAHASAN

Bab ini membahas tentang hasil dan pembahasan dari perancangan data mining Algoritma K- Means, K- Medoids dan pengujian Shilhoutte dan SSE , serta output yang dihasilkan berdasarkan analisis perancangan dan implementasi yang telah dilakukan pada proses data mining dengan metode yang digunakan pada Perhitungan Data Mining.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran berdasarkan pembahasan pada bab-bab sebelumnya