



Penentuan Tingkat Kekumuhan Permukiman Kumuh Kota Palembang Dengan Metode *Algoritma K-Means Clustering* Dan *Algoritma ID3*

Endah Puspita Sari¹, Lastris Widya Astuti², Imelda Saluza³, Faradillah⁴, Rini Yunita⁵

^{1,2} Program Studi Teknik Informatika, Universitas Indo Global Mandiri, Palembang, 30113, Indonesia, ³ Program Studi Manajemen Informatika, Universitas Indo Global Mandiri, Palembang, 30113, Indonesia, ⁴ Program Studi Sistem Informasi, Universitas Indo Global Mandiri, Palembang, 30113, Indonesia, ⁵ Program Studi Teknik Komputer, Sekolah Tinggi Teknologi, Payakumbuh, 26227, Indonesia.

Email : ¹endah_puspita91@yahoo.com, ²lastriwidya@uigm.ac.id, ³imeldasaluza@uigm.ac.id, ⁴faradillah.hakim@uigm.ac.id, ⁵riniyunita121n1@gmail.com

INFORMASI ARTIKEL

Diterima Redaksi: 7 Mei 2021

Revisi Akhir: 25 Mei 2021

Diterbitkan Online: 30 Mei 2021

KATA KUNCI

Keywords: Confusion Matrix, ID3 Algorithm, K-Means Clustering Algorithm, Slums

ABSTRACT

Slum settlers are a condition of uninhabitable settlements. Slums are divided into 4 levels, namely : high slums, medium, light and not slum. To produce these four levels, method is used namely K-Means Clustering Algorithm and the ID3 Algorithm is used to give priority with pre determined attributes then accumulated with the results of clustering classified as slum level. The accuracy test is performed by using the confusion matrix method, where the data results from the K-Means Clustering method compared to the baseline data. The results obtained from the accuracy with confusion matrix is 0,70%, which means the level of truth (accuracy) between the result of the baseline data with research data is 70%.

1. PENDAHULUAN

Permukiman kumuh merupakan keadaan lingkungan hunian dengan kualitas yang sangat tidak layak huni. Salah satu cara yang dilakukan untuk menangani kumuh adalah dengan mencegah dan meningkatkan kualitas perumahan dan permukiman kumuh guna memperbaiki mutu kehidupan dan penghidupan masyarakat. Pemerintah dalam hal ini Kementerian Pekerjaan Umum dan Perumahan Rakyat, Direktorat Jendral Cipta Karya bekerja sama dengan konsultan individual Program Kota Tanpa Kumuh (KOTAKU) kota Palembang akan melakukan identifikasi permukiman berdasarkan tingkatannya. Pada Peraturan Menteri PUPR No. 02/PRT/M/2016 tentang "Peningkatan Kualitas Terhadap Perumahan Kumuh dan Permukiman Kumuh" telah menentukan kriteria tingkat permukiman kumuh tersebut dengan menggunakan 7 indikator 19 parameter. Indikator-indikator tersebutlah yang akan menentukan kawasan *deliniasi* kumuh tersebut termasuk dalam tingkat kumuh berat, kumuh sedang, kumuh ringan, dan tidak kumuh.

Survey lapangan dilakukan untuk mendapatkan data-data yang berisi tentang 7 indikator 19 parameter. Data hasil *survey* akan membantu menentukan tingkat kekumuhan suatu kawasan, data tersebut diolah menjadi data per Kelurahan yang kemudian data tersebut dikelompok-kelompokkan atau yang biasa disebut dengan *Clustering*/klasterisasi.

Jurnal penelitian sebelumnya tentang klasterisasi oleh Atthina *et al* (2014) yang berjudul Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode *K-Means*, pada penelitian ini dapat dilihat penggunaan metode *K-Means* untuk mengklaster atau mengelompokkan kabupaten-kabupaten di Provinsi Jawa Tengah dan DIY berdasarkan kemiripan nilai lima buah ukuran indikator mortalitas derajat kesehatan daerah. *Data mining* dilihat dari sisi teknik pengolahan data menyediakan sejumlah *algoritma* yang dapat digunakan untuk menggali informasi tersembunyi dari kumpulan data yang multidimensi. Salah satu *algoritma data mining* yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah *algoritma K-Means Clustering*. *Algoritma* ini bekerja

dengan cara membagi data dalam sejumlah kluster untuk dianalisis faktor kesamaan (*similarity*) maupun ketidaksamaan (*dissimilarity*) yang melekat pada kumpulan data tersebut, dan kemudian dianalisis pola keterhubungan antar data.

Jurnal penelitian sebelumnya tentang *entropy* dengan metode *algoritma ID3* oleh Rumagit *et al* (2016) yang berjudul Penerapan Metode *ID3* terhadap Perancangan Sistem Penentuan Penerima Bantuan Sosial Pemugaran RTLH kota Salatiga, menghasilkan sistem pendukung keputusan dengan menggunakan *algoritma ID3* yang dapat digunakan oleh Badan Pemberdayaan Masyarakat, Perempuan, Keluarga Berencana dan Ketahanan Pangan (Bapermasper, KB & KP) untuk menentukan kelayakan penerima bantuan dalam perancangan sistem penerima Bansos RTLH. Metode *algoritma ID3* menjadi alternatif yang digunakan dalam penelitian ini karena *rule* lewat *decision tree* yang dihasilkannya mudah dimengerti dibandingkan dengan metode-metode klasifikasi lainnya, selain itu dengan kompleksitas data yang ada serta keragaman atribut didalamnya, membuat metode *algoritma ID3* menjadi pilihan yang terbaik dalam membantu penentuan keputusan. *Algoritma ID3* cocok untuk data-data yang memiliki tingkat variasi yang komplit.

Perhitungan dalam penentuan tingkat kumuh pada data *baseline* yang ada saat ini masih rumit, maka pada penelitian ini dibuat secara sederhana dan dapat melakukan prioritas kumuh untuk mempermudah dalam penanggulangan dini di Kelurahan. Metode klusterisasi dan metode *decision tree* yang digunakan memiliki kelebihan untuk diterapkan pada penelitian yang akan dibuat, dimana metode *algoritma K-Means Clustering* digunakan untuk mengelompokkan tingkat kumuh, sementara metode *algoritma ID3* untuk menentukan prioritas dari tingkat kumuh sebagai penanggulangan kumuh dini.

Berdasarkan latar belakang tersebut maka peneliti mencoba melakukan penelitian dengan judul : Penentuan Tingkat Kekumuhan Permukiman Kumuh kota Palembang dengan metode *Algoritma K-Means Clustering* dan *Algoritma ID3*. Diharapkan dengan adanya penelitian ini nantinya dapat membantu dalam menentukan prioritas penanggulangan kumuh suatu kelurahan.

2. TINJAUAN PUSTAKA

Model Analisis

A. Landasan Teori

Bagian ini memuat data teori yang diperlukan dan digunakan dalam penentuan tingkat kekumuhan permukiman kumuh kota Palembang dengan metode *Algoritma K-Means Clustering* dan *Algoritma ID3*.

1. Permukiman

Pengertian dasar permukiman dalam Undang-Undang No. 1 tahun 2011 adalah bagian dari lingkungan hunian yang terdiri atas lebih dari satu satuan perumahan yang mempunyai prasarana, sarana, utilitas umum, serta mempunyai penunjang kegiatan fungsi lain di kawasan perkotaan atau kawasan perdesaan.

2. Pengertian Kumuh

Kumuh adalah kesan atau gambaran secara umum tentang sikap dan tingkat laku yang rendah dilihat dari

standar hidup dan penghasilan kelas menengah. Dengan kata lain, kumuh dapat diartikan sebagai tanda atau cap yang diberikan golongan atas yang sudah mapan kepada golongan bawah yang belum mapan (Handryant, 2012).

Menurut Undang-Undang No. 1 Pasal 1 Tahun 2011 tentang Perumahan dan Kawasan Permukiman tidak layak huni karena ketidakteraturan bangunan, tingkat kepadatan bangunan yang tinggi, dan kualitas bangunan serta sarana dan prasarana yang tidak memenuhi syarat.

3. Clustering

Clustering atau klusterisasi adalah suatu teknik atau metode untuk mengelompokkan data. Menurut Tan (2006) *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum

Syarat Clustering

Menurut Han & Kamber (2012) syarat sekaligus tantangan yang harus dipenuhi oleh suatu *algoritma clustering* adalah :

1. Skalabilitas

Suatu metode *clustering* harus mampu menangani data dalam jumlah yang besar. Saat ini data dalam jumlah besar sudah sangat umum digunakan dalam berbagai bidang, misalnya saja suatu *database*. Tidak hanya berisi ratusan objek, suatu *database* dengan ukuran besar bahkan berisi lebih dari jutaan objek.

2. Kemampuan analisa beragam bentuk data

Algoritma klusterisasi harus mampu diimplementasikan pada berbagai macam bentuk data seperti data nominal, ordinal maupun gabungannya.

3. Menemukan *cluster* dengan bentuk yang tidak terduga

Banyak *algoritma clustering* yang menggunakan metode *Euclidean* atau *Manhattan* yang hasilnya berbentuk bulat. Padahal hasil *clustering* dapat berbentuk aneh dan tidak sama antara satu dengan yang lain. Karenanya dibutuhkan kemampuan untuk menganalisa *cluster* dengan bentuk apapun pada suatu *algoritma clustering*.

4. Kemampuan untuk dapat menangani *noise*

Data tidak selalu dalam keadaan baik. Ada kalanya terdapat data yang rusak, tidak dimengerti atau hilang. Karena sistem inilah, suatu *algoritma clustering* dituntut untuk mampu menangani data yang rusak.

5. Sensitifitas terhadap perubahan input

Perubahan atau penambahan data pada input dapat menyebabkan terjadi perubahan pada *cluster* yang telah ada bahkan bisa menyebabkan perubahan yang mencolok apabila menggunakan *algoritma clustering* yang memiliki tingkat sensitifitas rendah.

6. Mampu melakukan *clustering* untuk data dimensi tinggi

Suatu kelompok data dapat berisi banyak dimensi ataupun atribut. Untuk itu diperlukan *algoritma clustering* yang mampu menangani data dengan dimensi yang jumlahnya tidak sedikit.

7. Interpretasi dan kegunaan

Hasil dari *clustering* harus dapat diinterpretasikan dengan berguna.

3. METODE PENELITIAN

1. *Algoritma K-Means Clustering*

K-Means merupakan *algoritma clustering* yang berulang-ulang. *Algoritma K-Means* dimulai dengan pemilihan secara acak *K*, *K* disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai *K* secara *random*, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau “*means*”. Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidean* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (*distance space*). Beberapa cara yang telah diimplementasikan dalam menghitung jarak (*distance*) antara data dan *centroid* termasuk diantaranya *L₁* (*Manhattan/City Block distance space*), *L₂* (*Euclidean distance space*), dan *L_p* (*Minkowski distance space*) (Zainul et al., 2016).

Jarak antara dua titik *X₁* dan *X₂* pada *Manhattan/City Block Distance Space* dihitung dengan menggunakan rumus sebagai berikut :

$$D_{L1}(x_j, c_j) = \|x_j - c_j\|_1 = \sum_{j=1}^p |x_j - c_j| \quad (1)$$

Dimana :

- D_L* : Jarak antara dua titik
- p* : Banyak data
- c* : *Centroid*
- x* : Data
- | . |* : Nilai absolut

Sedangkan untuk *L₂* (*Euclidean distance space*), jarak antara dua titik dihitung menggunakan rumus sebagai berikut :

$$D_{L2}(x_j, c_j) = \|x_j - c_j\|_2 = \sqrt{\sum_{j=1}^p (x_j - c_j)^2} \quad (2)$$

Secara umum *distance space* yang sering digunakan adalah *Manhattan* dan *Euclidean*. *Euclidean* sering digunakan karena perhitungan jarak dalam *distance space* ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan. Sedangkan *Manhattan* sering digunakan karena kemampuan dalam mendeteksi keadaan khusus seperti keberadaan *outliers* dengan lebih baik.

Secara umum, langkah-langkah penyelesaian *algoritma K-Means Clustering* adalah sebagai berikut :

1. Tentukan banyaknya klaster (*k*).
2. Tentukan *centroid*. Jika *centroid* nilainya berubah, hitung jarak data dari *centroid*. Jika tidak, maka proses selesai.
3. Mengelompokkan data berdasarkan jarak yang terdekat.

2. *Decision Tree* (Pohon Keputusan)

Salah satu metode dalam klasifikasi *data mining* adalah *Decision Tree* (pohon keputusan). Klasifikasi tersebut membagi data ke kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Metode *Decision Tree* dapat

juga dikatakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia (Kusrini & Lutfi, 2009).

Konsep dasar *Decision Tree* adalah mengubah data menjadi pohon keputusan dan membuat aturan-aturan (*rule*), dimana setiap *node* merepresentasikan nilai dari atribut, cabangnya merepresentasikan kelas. *Node* yang paling atas dari *Decision Tree* disebut sebagai *root*. Ada banyak *algoritma* dari *Decision Tree*, salah satu diantaranya adalah *Iterative Dichotomizer Three* (*ID3*) (Utama et al., 2014).

3. *Algoritma ID3*

Algoritma ID3 adalah suatu metode induksi aturan yang digunakan untuk menghasilkan konsep atau model dari suatu kumpulan data. *ID3* diperkenalkan pertama kali oleh *Quinlan* (1979). *ID3* dikembangkan atas dasar sistem pembelajaran konsep (*Concept Learning System*) dari *Hut et al*, tujuan dari sistem pembelajaran konsep adalah untuk menghasilkan suatu pohon aturan yang mampu mengklasifikasi suatu objek (Manongga, 2005).

Algoritma ID3 memiliki cara kerja sebagai berikut :

- a. Ambil semua atribut yang tidak terpakai dan hitung entropinya yang berhubungan dengan *test sample*.
- b. Pilih atribut dimana nilai entropinya minimum.
- c. Buat simpul yang berisi atribut tersebut.

Untuk menghitung *ID3*, maka harus mencari nilai *entropy* dan *information gain*-nya dapat dilihat pada rumus dibawah ini :

Rumus Entropy :

$$Entropy(S) = -p_a \log_2 p_a - p_b \log_2 p_b \quad (3)$$

Keterangan :

- S* : Ruang (data) sampel yang digunakan untuk *training*
- p_a* : Probabilitas/jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu
- p_b* : Probabilitas/jumlah yang bersolusi negatif (tidak mendukung) pada data sampel untuk kriteria tertentu

Rumus Information Gain :

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{S} Entropy(S_v) \quad (4)$$

Keterangan :

- A* : Atribut
- V* : Menyatakan suatu nilai yang mungkin untuk atribut
- |S_v|* : Jumlah sampel untuk nilai *v*
- |S|* : Jumlah seluruh sampel data
- Entropy(S_v)* : *Entropy* untuk sampel-sampel yang memiliki nilai *v*

Catatan :

Entropy(S) = 0, jika semua sampel pada *S* berada dalam kelas yang sama.

Entropy(S) = 1, jika jumlah sampel positif dan jumlah sampel negatif dalam *S* adalah sama.

$0 < Entropy(S) < 1$, jika jumlah sampel positif dan jumlah sampel negatif dalam S tidak sama.

Langkah - langkah dari pembentukan pohon keputusan dengan *Algoritma ID3* sebagai berikut :

1. Memasukkan data klasterisasi, ini adalah syarat yang paling penting dari pembentukan *algoritma ID3*. Jadi pola keputusan dari data klasterisasi yang akan menjadi acuan dalam pembentukan pohon.
2. Melakukan perhitungan *Entropy* dan *Information Gain*, ini adalah tahap yang digunakan untuk menemukan simpul akar dari atribut yang ada sampai pohon berhenti terbentuk.
3. Setelah pohon sudah dibentuk, dibuatlah sebuah *rule* yang akan menjadi aturan dalam pohon.

4. Pengujian Data

Pengujian data pada penelitian yang akan dilakukan ini menggunakan metode *Confusion Matrix*. Data yang telah dilakukan klasterisasi dengan metode *K-Means Clustering* akan dihitung akurasi dengan menggunakan *Confusion Matrix*.

5. *Confusion Matrix*

Menurut Han & Kamber (2006) *confusion Matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Evaluasi dengan *confusion matrix* menghasilkan nilai akurasi. Akurasi dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi.

Tabel 1. Model Confusion matrix

		Predicted Class	
		yes	no
Actual Class	Yes	TP	FN
	No	FP	TN

(Han & Kamber, 2006, *Data Mining Concepts & Techniques*)

- TP (*True Positive*) : Jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif
- FP (*False Positive*) : Jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif
- FN (*False Negative*) : Jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif
- TN (*True Negative*) : Jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif

Untuk menguji data dari hasil klasterisasi dapat menggunakan rumus Akurasi. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

Rumus untuk menghitung nilai akurasi adalah sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

3.1. *Penulisan Tabel (10pt, tebal)*

Nomor dan judul tabel ditulis diposisi tengah kolom (*center alignment*).Tabel dinomori dengan angka arab

sesuai dengan urutannya. Judul tabel ditulis dibagian atas tabel dengan cara *title case*, kecuali untuk kata sambung dan kata depan. Ukuran huruf untuk judul tabel dan isi tabel adalah 8 (delapan). Sisi paling luar tabel tidak boleh melebihi batas margin kolom. Jika ukuran tabel tidak cukup ditulis dalam lebar 1 (satu), maka penulisannya dapat melintasi 2 kolom. Jarak baris yang digunakan antara tabel dengan kalimat di atasnya dan dibawahnya adalah 1 (satu) baris kosong. Tabel wajib menggunakan *layout* sesuai dengan Tabel 1 tanpa menggunakan garis lurus/vertikal. Setiap tabel harus diacu dalam tulisan dengan disertai nomor tabel dan diawali dengan huruf besar, misalnya Tabel 1.

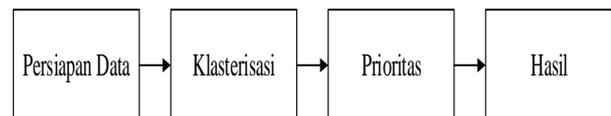
4. HASIL DAN PEMBAHASAN

A. *Teknik Pengumpulan Data*

Teknik pengumpulan data merupakan cara mengumpulkan data yang dibutuhkan untuk menjawab rumusan masalah penelitian. Metode penelitian didalam penelitian ini menggunakan metode penelitian kuantitatif, dengan teknik pengambilan data pada umumnya dilakukan dengan mengambil kelurahan yang menjadi kelurahan dampingan program KOTAKU. Data sekunder yang digunakan dalam penelitian ini merupakan data yang diambil oleh peneliti dari kantor Koordinator Kota 1 Program KOTAKU Palembang. Data yang diambil berupa data Baseline yang terdiri dari keseluruhan Kelurahan dampingan program KOTAKU yaitu berjumlah 106 Kelurahan.

B. *Deskripsi Umum Aplikasi*

Deskripsi umum aplikasi yang akan dirancang dapat dilihat pada **Gambar 1** :



Gambar 1. *Proses Umum Aplikasi*

Proses pada penelitian ini dimulai dari penginputan data *Baseline* yang diambil dari kantor Koordinator Kota KOTAKU kota Palembang. Tahap Persiapan Data dilakukan terlebih dahulu, setelah semua data sudah di persiapkan kemudian data diinput setelah itu data tersebut akan dilakukan proses Klasterisasi, setelah semua data telah terkelompok ke masing-masing klaster maka selanjutnya dilakukan proses prioritas. Klasterisasi dilakukan dengan menggunakan metode *Algoritma K-Means Clustering* yang hasilnya terbagi kedalam 4 kelompok yaitu Kumuh Berat, Kumuh Sedang, Kumuh Ringan dan Tidak Kumuh. Prioritas dilakukan setelah proses Klasterisasi didapatkan. Prioritas menggunakan metode *Algoritma ID3*. Prioritas dilakukan dengan melakukan penentuan nilai kelas *True* dan *False* terlebih dahulu untuk selanjutnya melakukan perhitungan *Algoritma ID3*. Hasil perhitungan ini akan dijadikan acuan para Fasilitator Kelurahan untuk menentukan prioritas penanganan kumuh.

C. Analisis Kebutuhan Aplikasi

Analisis sistem adalah proses memperhatikan atau meninjau sistem-sistem yang ada dan bertujuan untuk memperjelas tujuan dari pembuatan perangkat lunak yang sesuai dengan aplikasi yang akan dibangun. Sebelum perancangan dimulai diperlukan persiapan dengan bahan-bahan awal yang telah dibuat sebelumnya. Baik itu pengumpulan data yang akan diolah untuk diklasterisasi kemudian diprioritaskan.

1. Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional untuk mengetahui spesifikasi kebutuhan untuk aplikasi. Spesifikasi kebutuhan melibatkan kebutuhan perangkat lunak (*Software*), dan kebutuhan perangkat keras (*hardware*).

a) Kebutuhan Perangkat Lunak

Perangkat lunak merupakan suatu alat yang digunakan untuk membangun suatu aplikasi, berikut merupakan aplikasi yang digunakan untuk membuat penelitian ini :

1. Sistem Operasi *Windows 10*
2. *Microsoft Visio* untuk membuat *flowchart*
3. *CorelDRAW X7* untuk mendesign *background image*

b) Kebutuhan Perangkat Keras

Perangkat keras (*hardware*) adalah komponen dari sebuah komputer yang sifatnya bisa dilihat, diraba, dan dirasa secara langsung atau yang berbentuk nyata yang berfungsi untuk mendukung proses komputerisasi. Adapun komponen perangkat keras yang digunakan dalam membangun dan mendukung penelitian ini sebagai berikut :

1. *Processor AMD A4*
2. *Memory 4Gb*
3. *Hardisk 500 Gb*
4. *Mouse*

2. Kebutuhan Fungsional

Kebutuhan fungsional dimodelkan dengan *flowchart*.

a) *Flowchart* Penelitian

Flowchart Penelitian

Aliran data ini menjelaskan alur pada klasterisasi. Dalam klasterisasi ini memiliki beberapa inputan untuk diolah. Data yang diinput berdasarkan pada Peraturan Menteri PUPR No. 02/PRT/M/2016 berupa data *Baseline* yang memiliki kriteria penilaian 7 indikator kumuh antara lain : Keteraturan Bangunan, Aksesibilitas Lingkungan, Drainase Lingkungan, Pelayanan Air Minum, Pengelolaan Air Limbah, Pengelolaan Persampahan, dan Pengamanan Bahaya Kebakaran. Data yang telah tersedia tersebut diklasterisasi dengan menggunakan metode *Algoritma K-Means Clustering*. Kemudian data hasil klasterisasi ditentukan prioritasnya dengan menggunakan metode *Algoritma ID3* untuk menentukan prioritas penanggulangan wilayah kumuh dan akan ditampilkan hasil dari kedua metode tersebut pada tabel hasil.

D. Alur Proses (*Flowchart*)

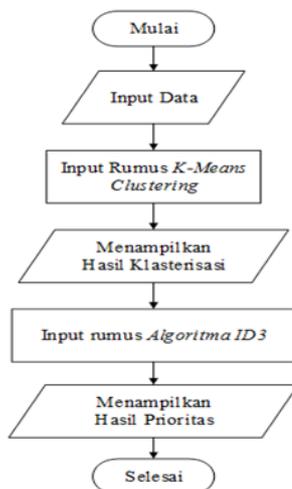
Alur proses keseluruhan penentuan tingkat kekumuhan permukiman kumuh kota Palembang dengan metode *Algoritma K-Means Clustering* dan *Algoritma ID3*. Dimana proses pertama dimulai dengan cara data yang telah disediakan dan dimasukkan ke proses klasterisasi dengan menggunakan metode *K-Means Clustering*. Hasil dari klasterisasi permukiman kumuh tersebut akan dilakukan prioritas untuk penanggulangan kumuh dengan *algoritma ID3*. Hasil akhir dari penelitian ini berupa prioritas penanggulangan kumuh

a) *Flowchart* Klasterisasi

Alur proses klasterisasi yang dilakukan pada penelitian ini adalah :

1. Penginputan data yang didapat dari tempat penelitian yaitu data *Baseline*.
2. Setelah data telah diinputkan, tentukan banyaknya *cluster* yang diperlukan.
3. Menentukan nilai *centroid* secara acak terlebih dahulu.

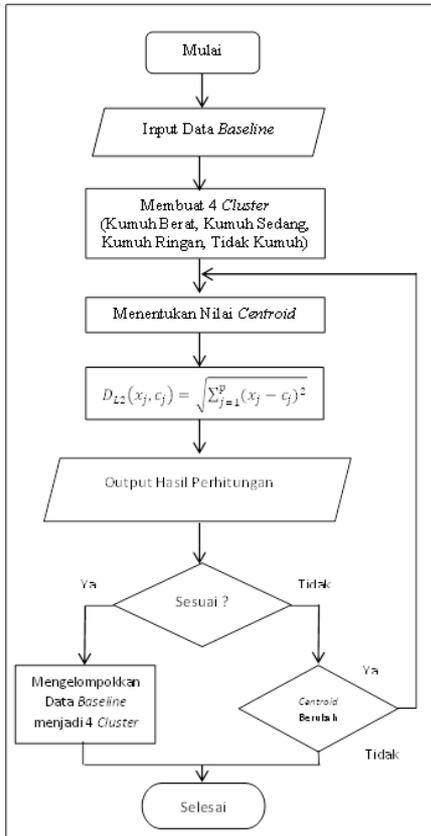
Gambar



2.

4. Setelah nilai *centroid* telah ditentukan, hitung jarak antara data *Baseline* dengan *centroid* dengan menggunakan rumus *K-Means Clustering*.

Perhitungan akan terus dilakukan jika nilai *centroid* mengalami perubahan, jika tidak mengalami perubahan maka perhitungan akan selesai dan akan mengelompokkan secara keseluruhan wilayah kelurahan tersebut menjadi 4 (empat) tingkat yaitu Kumuh Berat, Kumuh Sedang, Kumuh Ringan, dan Tidak Kumuh.



Gambar 3. Flowchart Klasterisasi

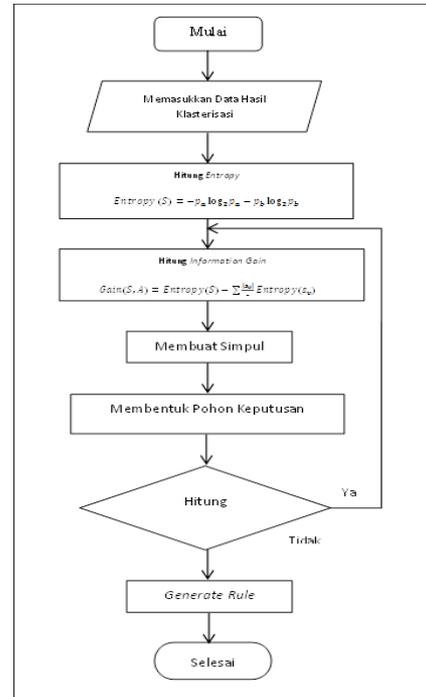
b) Flowchart Prioritas

Gambar 4. Flowchart Prioritas

Alur proses prioritas dengan *Algoritma ID3* dilakukan dengan cara sebagai berikut :

1. Ketika hasil klasterisasi telah didapat, diperlukan data baru untuk mempermudah menentukan atribut dalam mencari nilai *entropy* totalnya.
2. Hasil *entropy* total tersebut digunakan untuk mencari nilai *information gain* untuk masing-masing atributnya.
3. Nilai *information gain* yang terbesar akan membentuk simpul pada pohon keputusan.
4. Proses perhitungan terus dilakukan sampai semua atribut telah dihitung.

E. Pembahasan Implementasi



Pembahasan implementasi merupakan hasil atau gambaran dari implementasi penentuan tingkat kekumuhan permukiman kumuh kota Palembang dengan metode *Algoritma K-Means Clustering* dan *Algoritma ID3* pada sistem operasi *Windows 10* dengan menggunakan *MATLAB R2016b*.

a) Tampilan Layout

Tampilan *layout* adalah tampilan ketika *user* menjalankan aplikasi. Terdapat beberapa tampilan *layout*, seperti tampilan menu utama yang mana terdapat tampilan perhitungan, dan tampilan Informasi.

b) Tampilan Menu Utama

Toolbar yang digunakan pada *layout* Menu Utama ini adalah *Push Button*. *Push Button* yang digunakan ada 2 (dua) yaitu *Push Button* Perhitungan dan *Push Button* Informasi. *Push Button* Perhitungan digunakan untuk menampilkan tampilan perhitungan dari *Algoritma K-Means Clustering* dan *Algoritma ID3*, sedangkan *Push Button* Informasi digunakan untuk menampilkan tampilan informasi KOTAKU.



Gambar 5. Tampilan Menu Utama

c) Tampilan Perhitungan

Toolbar yang digunakan pada untuk membuat tampilan perhitungan ini adalah :

a. *Push Button*

Push Button yang digunakan ada 3 (tiga) yaitu Push Button Muat Data, Push Button Hasil, dan Push Button Tutup. Push Button Muat Data digunakan untuk menampilkan keseluruhan data baseline yang akan diuji. Push Button Hasil digunakan untuk menampilkan hasil perhitungan dari Algoritma K-Means Clustering dan algoritma ID3, sedangkan Push Button Tutup digunakan untuk menutup tampilan menu Perhitungan atau kembali ke menu utama.

b. Static Text

Static Text yang digunakan ada 2 (dua) yaitu static text Data Latih dan static text Centroids.

c. Table

Table yang digunakan ada 3 (tiga) yaitu Table Data Latih, Table Centroids dan Table Hasil. Table Data Latih digunakan untuk menampilkan data baseline yang akan diuji. Table Centroids digunakan untuk menampilkan data centroid dari Algoritma K-Means Clustering, dan Table Hasil digunakan untuk menampilkan perhitungan dari Algoritma K-Means Clustering untuk Tingkat Kumuh dan Algoritma ID3 untuk menentukan prioritas penanganan kumuh.



Gambar 6. Tampilan Perhitungan

d) Tampilan Informasi

Tampilan informasi berisi tentang informasi-informasi berupa nama sosial media program KOTAKU seperti nama Website resmi program KOTAKU, nama media sosial program KOTAKU dan nomor SMS Pengaduan (PPM). Adapun toolbar yang digunakan hanya Push Button Tutup. Push Button Tutup digunakan untuk menutup tampilan informasi atau kembali ke tampilan menu utama.



Gambar 7. Tampilan Informasi

F. Proses Klasterisasi

Proses Klasterisasi ini dilakukan dengan 106 Kelurahan dampingan yang ada di kota Palembang diproses dengan menggunakan metode K-Means Clustering. Data kelurahan tersebut akan diklaster menjadi 4 kategori tingkat Kumuh yang dapat dilihat pada Tabel 2 yaitu Tidak Kumuh, Kumuh Ringan, Kumuh Sedang dan Kumuh Berat. Berikut ini parameter clustering yang akan digunakan untuk Klasterisasi dengan K-Means Clustering dapat dilihat pada Tabel 3.

Tabel 2. Kelas Cluster

Tingkat Kumuh	Rentang Nilai	Cluster
Tidak Kumuh	0% - 25%	1
Kumuh Ringan	26% - 50%	2
Kumuh Sedang	51% - 75%	3
Kumuh Berat	76% - 100%	4

Tabel 3. Parameter K-Means Clustering

Jumlah Klaster (K)	4
Centroid	4 x 19
Maksimum Iterasi	100
Iterasi Awal	1

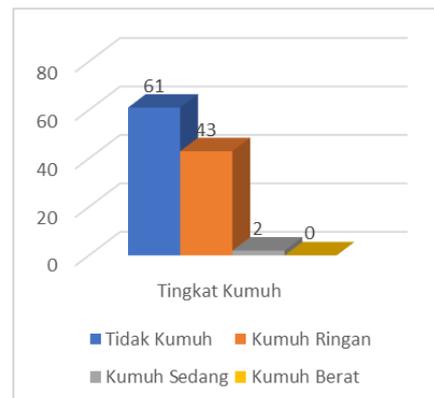
G. Proses Perhitungan

Perhitungan yang akan dilakukan pada penelitian ini menggunakan perhitungan klasterisasi dengan rumus dari algoritma K-Means Clustering dan perhitungan prioritas dengan rumus dari algoritma ID3. Perhitungan keakuratan data hasil dari data baseline dengan hasil dari data penelitian dilakukan dengan menggunakan rumus confusion matrix.

a) Proses Perhitungan Klasterisasi

Sebanyak 106 kelurahan tersebut akan dikelompokkan kedalam 4 cluster yaitu Cluster 1 (Tidak Kumuh), Cluster 2 (Kumuh Ringan), Cluster 3 (Kumuh Sedang), Cluster 4 (Kumuh Berat).

Hasil perhitungan dengan menggunakan Algoritma K-Means Clustering diperoleh ada sebanyak 61 kelurahan yang terklaster ke Cluster 1, sebanyak 43 kelurahan yang terklaster ke Cluster 2, sebanyak 2 kelurahan yang terklaster ke Cluster 3, dan sebanyak 0 kelurahan yang terklaster ke Cluster 4. Hasil dari perhitungan tersebut memberikan arti bahwa ada perbedaan hasil dengan data baseline.



b) Perhitungan Akurasi Klasterisasi dengan *Confusion Matrix*

Perbedaan hasil dari data *baseline* dengan data penelitian akan dilakukan pengujian perhitungan akurasi. Berikut ini merupakan perhitungan hasil akurasi klasterisasi dengan perhitungan *Confusion Matrix*.

Ketentuan untuk masing-masing nilai *True Positif*, *True Negatif*, *False Positif*, *False Negatif* dapat dijabarkan sebagai berikut :

True : Jika hasil data penelitian sesuai dengan data *baseline*

False : Jika hasil data penelitian tidak sesuai dengan data *baseline*

Positif : Jika hasil data penelitian terklaster ke Tidak Kumuh

Negatif : Jika hasil data penelitian terklaster ke Kumuh

Kesimpulan dari penentuan nilai *True*, & *False*, *Positif* dan *negatif* dapat dilihat pada **Tabel 5**.

Tabel 5. Hasil akurasi Klasterisasi dengan *Confusion Matrix*

	<i>True</i>	<i>False</i>
<i>Positif</i>	33	29
<i>Negatif</i>	41	3

Maka setelah didapatkan nilai *true* & *False*, *Positif* dan *Negatif* dapat dilakukan perhitungan akurasi dengan menggunakan rumus pada *Confusion Matrix* sebagai berikut :

Untuk Nilai Akurasi

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Akurasi = \frac{33+41}{33+41+29+3} \times 100\%$$

$$Akurasi = \frac{74}{106} \times 100\%$$

$$Akurasi = 70\%$$

c) Proses Prioritas

Proses prioritas dilakukan ketika telah didapatkan hasil dari klasterisasi. Kelurahan yang termasuk dalam kategori kumuh akan dilakukan penentuan prioritas dengan menggunakan metode *Algoritma ID3*. Sebelum dilakukan perhitungan dengan *Algoritma ID3*, terlebih dahulu dilakukan pembobotan untuk atribut-atribut *algoritma ID3* untuk mendapatkan kelas "*True*" dan kelas "*False*". Untuk menentukan nilai bobot diperlukan penilaian yang dapat dilihat pada **Tabel 6**.

Tabel 6. Kriteria Nilai Bobot

Kriteria Nilai	Bobot
70% - 100%	True
0% - 69%	False

Dengan 7 atribut yang digunakan untuk perhitungan metode *Algoritma ID3*. Ke-7 atribut tersebut didapatkan dari 7 indikator permukiman kumuh yang terdapat pada data *baseline*.

5. KESIMPULAN DAN SARAN

Kesimpulan yang diperoleh pada penelitian ini adalah dari hasil perhitungan dengan metode *algoritma K-Means Clustering* didapatkan jumlah klaster yang tergolong tidak kumuh sebanyak 61 kelurahan, yang tergolong kumuh ringan sebanyak 43 kelurahan, yang tergolong kumuh sedang sebanyak 2 kelurahan dan tidak ada kelurahan yang tergolong dalam kumuh berat.

Dengan nilai akurasi yang didapatkan dari perhitungan *confusion matrix*, maka tingkat ketepatan (kebenaran) hasil antara data *baseline* dengan data penelitian sebesar 70%.

Metode algoritma ID3 pada penelitian ini dapat digunakan untuk menentukan prioritas kumuh. Informasi yang diberikan berupa kelurahan yang termasuk dalam prioritas dan kelurahan yang termasuk dalam tidak prioritas. Hasil dari algoritma ID3 dapat digunakan sebagai bahan pertimbangan untuk penanganan kumuh dini diwilayah dampingan.

DAFTAR PUSTAKA

[1] Atthina, N., & Iswari, L. (2014). Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means.

[2] Han, J Kamber, M. (2006) *Data Mining Concepts & Techniques*. 2nd edn. San Fransisco: Elsevier.

[3] Handryant, A. N. (2012). Permukiman Kumuh, Sebuah Kegagalan Pemenuhan Aspek Permukiman Islami. *Journal of Islamic Architecture*, 1(3). <https://doi.org/10.18860/jia.v1i3.1774>

[4] Kusriani & Lutfi, E. . (2009). *Algoritma Data Mining*. Yogyakarta: ANDI.

[5] Manongga, D. (2005). *Teori dan Aplikasi Iterative Dichotomizer Three dalam Pembelajaran Mesin*. Salatiga: Universitas Kristen Satya Wacana.

[6] Rumagit, C. M., Fibriani, C., No, J. D., Salatiga, K., & Tengah, J. (2016). Penerapan Metode ID3 terhadap Perancangan Sistem Penentuan Penerima Bantuan Sosial Pemugaran RTLH Kota Salatiga, 02(03).

[7] Tan, P.-N. (2006). *Introduction to Data Mining*. Boston: Pearson Addison.

[8] Undang-Undang Nomor 1 tahun 2011

[9] Utama, T. D., Informatika, J., Maret, U. S., Sihwi, S. W., Informatika, J., Maret, U. S., ... Maret, U. S. (2014). Vol 3 . No 2 . Desember 2014 ISSN : 2301 – 7201 IMPLEMENTASI ALGORITMA ITERATIVE DICHOTOMISER 3 PADA, 3(2), 74–83.

[10] Z, Z. A., & Sarjono. (2016). Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means (Studi Kasus: Kantor Kecamatan Bahar Utara). *Z, Zainul Aras Sarjono*, 1(2), 159–170