

**ANALISIS *DATA MINING* CLUSTERING TENTANG TINGKAT
KESEJAHTERAAN DI DESA KARANG ANYAR
MENGUNAKAN METODE *K-MEANS***

SKRIPSI

Untuk Memenuhi Persyaratan Memperoleh Gelar Sarjana (S1)

Program Studi Sistem Informasi Fakultas Sains dan Teknologi

Universitas Labuhanbatu



OLEH:

NUR HALIMAH TAMBUNAN

2009100092

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS LABUHANBATU

RANTAUPRAPAT

2024

LEMBAR PENGESAHAN/PERSETUJUAN SKRIPSI

JUDUL : ANALISIS DATA MINING CLUSTERING TENTANG
TINGKAT KESEJAHTERAAN DI DESA KARANG ANYAR
MENGUNAKAN METODE K-MEANS
NAMA MAHASISWA : NUR HALIMAH TAMBUNAN
NPM : 2009100092
PROGRAM STUDI : SISTEM INFORMASI

Pada Tanggal : 19 Agustus 2024

PEMBIMBING I



MARNIS NASUTION, S.Kom., M.Kom
NIDN. 0130039001

PEMBIMBING II



RAHMA MUT'AH, S.Psi., M.Psi
NIDN. 0114068501

LEMBAR PENGESAHAN NASKAH SKRIPSI

JUDUL : ANALISIS DATA MINING CLUSTERING TENTANG
TINGKAT KESEJAHTERAAN DI DESA KARANG ANYAR
MENGUNAKAN METODE K-MEANS

NAMA MAHASISWA : NUR HALIMAH TAMBUNAN

NPM : 2009100092

PROGRAM STUDI : SISTEM INFORMASI

KONSENTRASI : DATA MINING DAN BIG DATA

Yang dipersiapkan dan disusun oleh

Telah Diuji Dan Dinyatakan Lulus Dalam Ujian Sarjana
Pada Tanggal 19 Agustus 2024

TIM PENGUJI

Penguji I (Ketua)

Nama : MARNIS NASUTION, S.Kom., M.Kom.
NIDN : 0130039001

Tanda Tangan



Penguji II (Anggota)

Nama : RAHMA MUTI'AH, S.Psi., M.Psi
NIDN : 0114068501




Penguji III (Anggota)

Nama : SYAIFUL ZUHRI HARAHAHAP, S.Kom., M.Kom
NIDN : 0113129103



Rantauprapat, 19 Agustus 2024

Ketua Program Studi
Sistem Informasi



BUDI ANTO BANGUN, S.Kom., M.Kom
NIDN. 0124047003

Diketahui Oleh:
Dekan Fakultas Sains dan Teknologi



Dr. IWAN PURNAMA, S.Kom., M.Kom
NIDN. 0112029202

SURAT PERNYATAAN

Nama : NUR HALIMAH TAMBUNAN
NPM : 2009100092
JUDUL : ANALISIS DATA MINING CLUSTERING TENTANG TINGKAT
KESEJAHTERAAN DI DESA KARANG ANYAR
MENGUNAKAN METODE K-MEANS

Dengan ini penulis menyatakan bahwa Skripsi ini disusun sebagai syarat untuk memperoleh gelar sarjana pada Program Studi Sistem Informasi Fakultas Sains dan Teknologi Rujukan dalam Skripsi ini telah penulis cantumkan sumbernya dengan benar dan sesuai dengan ketentuan yang berlaku.

Apabila kemudian hari bahwa pernyataan saya tidak benar, saya bersedia menanggung resiko dan siap diperkarakan sesuai dengan aturan berlaku. Demikian surat pernyataan ini buat dengan sesungguhnya.

Rantauprapat, 19 Agustus 2024

Yang Menyatakan



NUR HALIMAH TAMBUNAN
NPM. 2009100092

ABSTRAK

Analisis kluster tingkat kesejahteraan masyarakat memiliki tujuan utama untuk mengidentifikasi pola dalam data yang memberikan gambaran jelas mengenai sebaran kesejahteraan di suatu wilayah. Dengan menggunakan pendekatan ini, masyarakat dapat dikelompokkan berdasarkan karakteristik ekonomi, sosial, dan demografi yang mereka miliki. Hasil dari pengelompokan ini sangat berguna untuk menentukan wilayah-wilayah mana yang memerlukan perhatian lebih dalam hal pembangunan. Identifikasi pola tersebut tidak hanya memetakan keadaan kesejahteraan secara umum, tetapi juga membantu dalam merancang strategi yang lebih efektif dan efisien dalam menyalurkan program-program kesejahteraan yang ada. Dengan demikian, setiap individu dalam masyarakat diharapkan dapat menerima manfaat yang tepat sesuai dengan kebutuhan mereka. Dalam analisis ini, metode *K-Means* digunakan sebagai teknik klusterisasi untuk mengelompokkan data berdasarkan fitur-fitur yang serupa. Metode ini dipilih karena keefektifannya dalam mengelompokkan data dengan berbagai karakteristik yang berbeda. Dalam konteks tingkat kesejahteraan masyarakat, data yang dianalisis mencakup berbagai indikator penting seperti pendapatan, tingkat pendidikan, dan akses terhadap layanan kesehatan. *K-Means* kemudian membagi data tersebut ke dalam beberapa kluster yang masing-masing mewakili tingkat kesejahteraan yang berbeda-beda. Setiap kluster yang terbentuk menggambarkan sekelompok orang dengan karakteristik yang serupa, sehingga memudahkan identifikasi wilayah yang mungkin memerlukan intervensi lebih lanjut untuk meningkatkan kesejahteraan penduduknya. Hasil dari analisis kluster ini menunjukkan bahwa masyarakat dapat dibagi ke dalam beberapa kelompok berdasarkan tingkat kesejahteraan yang mereka miliki. Dalam analisis yang dilakukan, terdapat kluster-kluster tertentu yang menggambarkan masyarakat dengan tingkat kesejahteraan yang relatif tinggi, sementara kluster lainnya menunjukkan tingkat kesejahteraan yang lebih rendah. Dengan hasil ini, para pemangku kepentingan dapat menggunakan informasi yang dihasilkan untuk merancang strategi yang lebih terarah dalam mengatasi ketimpangan sosial dan ekonomi. Secara keseluruhan, analisis kluster ini memberikan keakuratan yang cukup tinggi, bahkan meskipun perhitungan dilakukan secara manual. Keakuratan ini memberikan keyakinan bahwa metode *K-Means* adalah alat yang efektif dalam mengungkap pola-pola tersembunyi dalam data kesejahteraan masyarakat. Dari 60 data sampel yang dianalisis, hasil clustering menunjukkan bahwa data tersebut terbagi ke dalam tiga cluster yang berbeda. Cluster C1 terdiri dari 12 sampel, yang mungkin mewakili kelompok dengan karakteristik atau pola yang cukup berbeda dari kelompok lainnya. Cluster C2 mencakup 18 sampel, yang menandakan bahwa karakteristik dalam kelompok ini sedikit lebih umum.

Kata kunci: Penemuan Pengetahuan dalam Basis Data (KDD); Penambangan Data; Metode *K-Means*; Pengelompokan;

ABSTRACT

The main objective of community welfare cluster analysis is to identify patterns in data that provide a clear picture of the distribution of welfare in a region. Using this approach, communities can be grouped based on their economic, social, and demographic characteristics. The results of this grouping are very useful for determining which areas require more attention in terms of development. Identification of these patterns not only maps the general welfare situation but also helps in designing more effective and efficient strategies in distributing existing welfare programs. Thus, every individual in society is expected to receive the right benefits according to their needs. In this analysis, the K-Means method is used as a clustering technique to group data based on similar features. This method was chosen because of its effectiveness in grouping data with various different characteristics. In the context of community welfare, the data analyzed includes various important indicators such as income, education level, and access to health services. K-Means then divides the data into several clusters, each of which represents a different level of welfare. Each cluster formed depicts a group of people with similar characteristics, making it easier to identify areas that may require further intervention to improve the welfare of its population. The results of this cluster analysis show that society can be divided into several groups based on the level of welfare they have. In the analysis conducted, there are certain clusters that depict people with relatively high levels of welfare, while other clusters show lower levels of welfare. With these results, stakeholders can use the information generated to design more targeted strategies to address social and economic inequality. Overall, this cluster analysis provides quite high accuracy, even though the calculations are done manually. This accuracy provides confidence that the K-Means method is an effective tool in revealing hidden patterns in community welfare data. Of the 60 sample data analyzed, the clustering results show that the data is divided into three different clusters. Cluster C1 consists of 12 samples, which may represent a group with characteristics or patterns that are quite different from other groups. Cluster C2 includes 18 samples, indicating that the characteristics in this group are slightly more common.

Keywords: Knowledge Discovery In Database (KDD); Data Mining; Metode K-Means; Clustering;

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita sehingga saya dapat menyelesaikan Laporan Skripsi dengan judul “Analisis *Data Mining* Clustering Tentang Tingkat Kesejahteraan di Desa Karang Anyar Menggunakan Metode *K-Means*”. Laporan Skripsi ini disusun sebagai salah satu syarat untuk mengerjakan tugas akhir pada program Studi Sistem Informasi Fakultas Sains Dan Teknologi universitas labuhanbatu.

Saya sebagai Penulis menyadari dalam penyusunan Laporan Skripsi ini tidak akan dapat terselesaikan tanpa bantuan dari berbagai pihak. Karena itu pada kesempatan ini saya ingin mengucapkan terima kasih kepada:

1. Bapak Pendiri Universitas Labuhanbatu, Alm Bapak Dr. H. Amarullah Nasution, S.E., M.B.A.
2. Bapak Ketua Yayasan Universitas Labuhanbatu, Bapak Halomoan Nasution, S.H., M.H.
3. Bapak Rektor Universitas Labuhanbatu, Bapak Assoc, Prof. Ade Parlaungan Nasution, Ph.D
4. Bapak Dr. Iwan Purnama, S.Kom., M.Kom selaku Dekan Fakultas Sains Dan Teknologi Universitas Labuhanbatu.
5. Bapak Budianto Bangun, S.Kom., M.Kom selaku Kepala Program Studi Sistem Informasi.
6. Ibu Marnis Nasution, S.Kom., M.Kom selaku Dosen Pembimbing 1 (Satu).
7. Ibu Rahma Muti'ah, S.Psi., M.Psi selaku Dosen Pembimbing 2 (Dua)

8. Bapak Syaiful Zuhri Harahap, S.Kom., M.Kom selaku dosen Penguji saya

Saya menyadari Laporan Skripsi ini tidak luput dari berbagai kekurangan. Penulisan mengharapkan saran dan kritik untuk perbaikannya sehingga Laporan Skripsi ini dapat memberikan manfaat bagi banyak orang-orang dalam bidang komputer.

Rantauprapat, 2024
Penulis



Nur Halimah Tambunan
NIM. 2009100092

DAFTAR ISI

| | |
|---|------|
| HALAMAN JUDUL..... | i |
| ABSTRAK | ii |
| ABSTRACT | iii |
| KATA PENGANTAR..... | iv |
| DAFTAR ISI | vi |
| DAFTAR TABEL..... | viii |
| DAFTAR GAMBAR | ix |
| BAB I | 1 |
| PENDAHULUAN | 1 |
| 1.1.Latar Belakang Masalah | 1 |
| 1.2.Perumusan Masalah..... | 5 |
| 1.3.Ruang Lingkup Masalah | 6 |
| 1.4.Tujuan dan Manfaat Penelitian..... | 6 |
| 1.4.1.Tujuan Penelitian | 6 |
| 1.4.2.Manfaat Penelitian | 7 |
| 1.5.Tinjauan Umum Objek Penelitian | 7 |
| 1.6.Sistematika Penulisan..... | 8 |
| BAB II..... | 10 |
| LANDASAN TEORI | 10 |
| 2.1. <i>Data Mining</i> | 10 |
| 2.2. <i>Knowledge Discovery In Database (KDD)</i> | 11 |
| 2.3. <i>Algoritma K-Means Clustering</i> | 13 |
| 2.4.1.Uji Performa | 18 |
| 2.4.Alat Bantu Pemrograman/ <i>Tools</i> Pendukung..... | 19 |
| 2.4.1.Aplikasi <i>Orange</i> | 19 |
| 2.5.Metodologi Penelitian | 20 |
| 2.6.1.Penelitian Terdahulu | 21 |
| 2.6.Kerangka Penelitian | 23 |
| 2.7.Tingkat Kesejahteraan | 33 |
| BAB III | 35 |
| ANALISIS DAN PERANCANGAN..... | 35 |
| 3.1.Arsitektur Sistem..... | 35 |
| 3.2.Desain Aktifitas Sistem | 36 |
| 3.2.1.Pengolahan Data pada Metode <i>K-Means</i> | 36 |
| BAB IV | 50 |
| HASIL DAN PEMBAHASAN | 50 |
| 4.1.Seleksi Data..... | 50 |
| 4.2.Perancangan Model <i>Clustering</i> | 53 |
| 4.3.Hasil <i>Cluster</i> | 59 |
| BAB V..... | 63 |
| KESIMPULAN DAN SARAN..... | 63 |
| 5.1.Kesimpulan..... | 63 |
| 5.2.Saran..... | 63 |
| DAFTAR PUSTAKA | 65 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1. Confusion Matrix Metode Naïve Bayes | 18 |
| Tabel 2. 2. Penelitian Terdahulu | 21 |
| Tabel 3. 1. Data Sampel Penelitian | 37 |
| Tabel 3. 2. Cluster 1, 2, dan 3 Iterasi 1 | 40 |
| Tabel 3. 3. CLuster 4 Iterasi 1 | 41 |
| Tabel 3. 4. Update Centroid | 41 |
| Tabel 3. 5. Cluster 5 Iterasi 1 | 42 |
| Tabel 3. 6. Update Centroid | 42 |
| Tabel 3. 7. Cluster 6 Iterasi 1 | 42 |
| Tabel 3. 8. Update Centroid | 43 |
| Tabel 3. 9. Cluster 7 Iterasi 1 | 43 |
| Tabel 3. 10. Update Centroid | 43 |
| Tabel 3. 11. Cluster 8 Iterasi 1 | 44 |
| Tabel 3. 12. Update Centroid | 44 |
| Tabel 3. 13. Cluster 9 Iterasi 1 | 45 |
| Tabel 3. 14. Update Centroid | 45 |
| Tabel 3. 15. Cluster 10 Iterasi 1 | 45 |
| Tabel 3. 16. Hasil Clustering Iterasi 1 | 46 |
| Tabel 3. 17. Hasil Clustering Iterasi 2 | 49 |
| Tabel 4. 1. Data Sampel Penelitian | 50 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1. Centroid-based Clustering..... | 14 |
| Gambar 2. 2. Density-based Clustering | 15 |
| Gambar 2. 3. Distributin-based Clustering | 16 |
| Gambar 2. 4. Hierachical Clustering..... | 16 |
| Gambar 2. 5. Kerangka Kerja Penelitian | 23 |
| Gambar 3. 1. Kerangka Kerja Penelitian | 35 |
| Gambar 4. 1. Perancangan Model Clustering | 58 |
| Gambar 4. 2. Widget File | 53 |
| Gambar 4. 3. Widget Data Tabel untuk Data | 54 |
| Gambar 4. 4. Widget Metode <i>K-Means</i> | 54 |
| Gambar 4. 5. Widget Predictions | 55 |
| Gambar 4. 6. Widget Scatter Plot..... | 56 |
| Gambar 4. 7. Widget Data Tabel untuk Hasil | 56 |
| Gambar 4. 8. Widget Save File | 57 |