

SKRIPSI

**DATA MINING TUBERKULOSIS PADA ANAK DENGAN
K-MEANS CLUSTERING**

Disusun dan diajukan oleh

MUHAMMAD IRZAM KASYFILLAH

D121171308



DEPARTEMEN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

LEMBAR PENGESAHAN SKRIPSI
DATA MINING TUBERKULOSIS PADA ANAK DENGAN K-MEANS
CLUSTERING

Disusun dan diajukan oleh
MUHAMMAD IRZAM KASYFILLAH
D121171308

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 08 September 2022 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Dr. Indrabayu, S.T., M.T., M.Bus.Sys.
Nip. 197507162002121004



Elly Warni, S.T., M.T.
Nip. 198202162008122001

Ketua Program Studi,



Dr. Indrabayu, S.T., M.T., M.Bus.Sys.
Nip. 19750716 200212 1 004

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Irzam Kasyfillah

NIM : D121171308

Departemen : Teknik Informatika

Jenjang : SI

Menyatakan dengan ini karya tulisan saya berjudul:

DATA MINING TUBERKULOSIS PADA ANAK DENGAN K-MEANS CLUSTERING

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilalihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 08 September 2022

Yang menyatakan,



Muhammad Irzam Kasyfillah

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan rahmat dan karunia-Nya, sehingga dapat menyelesaikan tugas akhir dengan judul “Data Mining Tuberkulosis pada Anak dengan K-Means Clustering” sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 di Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Penulis menyadari bahwa penyusunan dan penulisan tugas akhir ini tidak dapat terselesaikan dengan baik tanpa adanya bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih banyak kepada:

1. Kedua orang tua penulis, Bapak Deddy Amrand dan Ibu Rahmawati Amin yang selalu memberikan dukungan, doa, dan semangat yang tiada hentinya, serta selalu sabar dalam mendidik penulis sejak kecil.
2. Bapak Dr. Indrabayu, S.T., M.T., M.Bus.sys., selaku pembimbing I dan Ibu Elly Warni, S.T., M.T. selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
3. Bapak Dr. Indrabayu, S.T., M.T., M.Bus.sys., selaku ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.
4. Dr. dr. Bob Wahyudin, Sp.A(K), CIMI yang senantiasa memberikan nasihat dan masukan, kak Ophy Sangbara yang membantu dalam pengumpulan data dan seluruh staf di Klinik Pediatrica Husada yang telah membantu penulis selama penelitian, serta Muhammad Alfarabi Alif Putra sebagai tim Tuberkulosis Anak.
5. Segenap keluarga AIMP *Research Group* Universitas Hasanuddin yang telah memberikan begitu banyak bantuan selama penelitian, pengambilan data dan diskusi progres penyusunan tugas akhir serta memberikan semangat di masa-masa sulit.
6. Muhammad Zulfahmi Sadrah, Muhammad Ilham Askari, Khaerul Hidayat, Muhammad Bishram, Muhammad Fadhil, dan Fauzan Alif Anwar serta seluruh

anggota Recognizer yang telah membantu penulis sejak awal perkuliahan dan selalu membantu dalam penyelesaian tugas akhir.

7. Segenap staf dan dosen Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir.
8. Serta pihak-pihak lain yang tidak disebutkan dan tanpa sadar telah menjadi inspirasi dan membantu penulis dalam menyelesaikan tugas akhir.

Penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan serta jasa dari semua pihak yang telah banyak membantu penulis dalam menyelesaikan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberi manfaat bagi para pembaca dan semua pihak. Aamiin.

Makassar, September 2022

Penulis

ABSTRAK

Tuberkulosis (TB) adalah penyakit menular yang merupakan penyebab utama kesehatan yang buruk, salah satu dari 10 penyebab kematian teratas di dunia dan penyebab utama kematian dari agen infeksi tunggal (peringkat di atas HIV/AIDS). Masalah yang dihadapi pada TB anak adalah kesulitan diagnosis, pengobatan dan belum optimalnya program pencegahan. Berdasarkan data yang diperoleh melalui pengisian kuesioner oleh orang tua pasien tuberkulosis anak Klinik Pediatrica Husada, terdapat beberapa variabel yang berhubungan dengan tuberkulosis anak. Penelitian ini bertujuan untuk membuat pengelompokan data tuberkulosis pada anak per kecamatan di Kota Makassar dan Kabupaten Gowa. Terdapat 18 parameter yang digunakan, yaitu jumlah TB, usia, persentase status gizi, persentase pendapatan orang tua, persentase luas rumah, persentase telah BCG, ASI eksklusif, TB serumah, anak menderita diabetes, dan keluarga menderita diabetes. Metode yang digunakan dalam *clustering* adalah *k-means clustering*. Hasil *clustering* yang terbentuk yaitu sebanyak 20 *cluster*. Dimana tiap kecamatan dapat termasuk dalam 1 *cluster* atau lebih yang menunjukkan *cluster* yang paling berpengaruh terhadap kasus TB anak di kecamatan tersebut. Hasil *clustering* ditampilkan dalam bentuk visualisasi webGIS yang menampilkan informasi *cluster* tiap-tiap kecamatan di Kota Makassar dan Kabupaten Gowa.

Kata kunci: *tuberkulosis anak, data mining, k-means clustering, webGIS*

DAFTAR ISI

KATA PENGANTAR	ii
ABSTRAK	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	xi
BAB 1.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2.....	6
TINJAUAN PUSTAKA.....	6
2.1 Tuberkulosis.....	6
2.2 Anak	7
2.3 Data Mining	8
2.4 Clustering	10
2.5 K-Means Clustering	11
2.6 Silhouette Coefficient.....	13
2.7 Sum of Squared Error (SSE)	14
2.8 Web GIS.....	15

2.9 System Usability Scale.....	16
BAB 3.....	18
METODOLOGI PENELITIAN.....	18
3.1 Tahapan Penelitian.....	18
3.2 Waktu dan Lokasi Penelitian.....	20
3.3 Instrumen Penelitian.....	20
3.4 Teknik Pengambilan Data.....	21
3.5 Perancangan Sistem.....	22
3.6 Clustering Algoritma K-Means.....	29
3.7 Pattern Evaluation.....	30
3.8 Visualisasi WebGIS.....	30
3.9 Evaluasi Usability.....	30
BAB 4.....	31
HASIL DAN PEMBAHASAN.....	31
4.1 Hasil Penelitian.....	31
4.1.1 Clustering berdasarkan Jumlah TB dan usia.....	31
4.1.2 Clustering berdasarkan Status Gizi Anak.....	38
4.1.3 Clustering berdasarkan Pendapatan Orang Tua.....	42
4.1.4 Clustering berdasarkan Luas Rumah.....	47
4.1.5 Clustering berdasarkan Persentase Telah BCG, TB Serumah, ASI Eksklusif, Diabetes Anak dan Diabetes Keluarga.....	53
4.2 Pembahasan.....	58
4.2.1 Interpretasi Hasil Clustering berdasarkan Jumlah TB dan Usia.....	58
4.2.2 Interpretasi Hasil Clustering berdasarkan Status Gizi Anak.....	59
4.2.3 Interpretasi Hasil Clustering berdasarkan Pendapatan Orang Tua....	59

4.2.4 Interpretasi Hasil Clustering berdasarkan Luas Rumah	60
4.2.5 Interpretasi Hasil Clustering berdasarkan Persentase Telah BCG, ASI Eksklusif, TB Serumah, Diabetes Anak dan Diabetes Keluarga.....	61
4.2.6 Observasi Hasil Clustering	62
4.2.7 Interpretasi Hasil Clustering untuk Setiap Kecamatan di Kota Makassar dan Kabupaten Gowa.....	64
4.2.8 Visualisasi WebGIS.....	71
4.2.9 Hasil Evaluasi Usability	76
BAB 5.....	78
KESIMPULAN DAN SARAN.....	78
5.1 Kesimpulan	78
5.2 Saran.....	78
DAFTAR PUSTAKA	80
LAMPIRAN.....	82

DAFTAR GAMBAR

Gambar 1.1 Angka Notifikasi Semua Kasus Tuberkulosis per 100.000 Penduduk Menurut Provinsi Tahun 2018	1
Gambar 2.1 Proses Knowledge Discovery in Database (Tan et al, 2019).....	9
Gambar 2.2 Tahap Data Mining (Han et al., 2011).....	9
Gambar 2.3 Ilustrasi 3 cara yang berbeda untuk melakukan clustering.....	11
Gambar 2.4 Flowchart Algoritma K-Means.....	13
Gambar 2.5 Interpretasi skor SUS (Varizi et al, 2016)	17
Gambar 3.1 Tahapan Penelitian.....	18
Gambar 3.2 Rancangan Sistem.....	22
Gambar 3.3 Sampel Data Tuberkulosis Anak	23
Gambar 3.4 Sampel Data Cleaning	24
Gambar 3.5 Sampel transformasi data daftar penyakit.....	25
Gambar 3.6 Daftar penyakit opname, penyakit orang serumah dan persentasenya	26
Gambar 3.7 Sampel transformasi data status gizi.....	27
Gambar 3.8 Data tuberkulosis per kecamatan	28
Gambar 3.9 Sampel data setelah normalisasi	28
Gambar 3.10 Jumlah TB per kabupaten/kota	29
Gambar 4.1 Grafik hubungan k terhadap SSE berdasarkan jumlah tb dan usia..	32
Gambar 4.2 Grafik hubungan k terhadap silhouette berdasarkan jumlah tb dan usia	32
Gambar 4.3 Centroid awal algoritma k-means.....	33
Gambar 4.4 Jarak tiap data terhadap centroid pada iterasi pertama	34
Gambar 4.5 Sampel penentuan cluster pada iterasi pertama	35
Gambar 4.6 Nilai centroid yang baru setelah iterasi pertama.....	35
Gambar 4.7 Nilai centroid yang dihasilkan tidak berubah pada iterasi ke-5.....	36
Gambar 4.8 Sampel penentuan cluster pada iterasi ke-5.....	36
Gambar 4.9 Daftar kecamatan pada cluster 1 berdasarkan jumlah TB dan usia .	37
Gambar 4.10 Daftar kecamatan pada cluster 2 berdasarkan jumlah TB dan usia	37

Gambar 4.11 Daftar kecamatan pada cluster 3 berdasarkan jumlah TB dan usia 38	
Gambar 4.12 Grafik hubungan k terhadap SSE berdasarkan status gizi anak	39
Gambar 4.13 Grafik hubungan k terhadap silhouette berdasarkan status gizi anak	39
Gambar 4.14 Daftar kecamatan pada cluster 1 berdasarkan status gizi	40
Gambar 4.15 Daftar kecamatan pada cluster 2 berdasarkan status gizi	41
Gambar 4.16 Daftar kecamatan pada cluster 3 berdasarkan status gizi	41
Gambar 4.17 Grafik hubungan k terhadap SSE berdasarkan pendapatan orang tua	42
Gambar 4.18 Grafik hubungan k terhadap silhouette berdasarkan pendapatan orang tua.....	43
Gambar 4.19 Daftar kecamatan pada cluster 1 berdasarkan pendapatan orang tua	44
Gambar 4.20 Daftar kecamatan pada cluster 2 berdasarkan pendapatan orang tua	45
Gambar 4.21 Daftar kecamatan pada cluster 3 berdasarkan pendapatan orang tua	45
Gambar 4.22 Daftar kecamatan pada cluster 4 berdasarkan pendapatan orang tua	46
Gambar 4.23 Daftar kecamatan pada cluster 5 berdasarkan pendapatan orang tua	46
Gambar 4.24 Grafik hubungan k terhadap SSE berdasarkan luas rumah	47
Gambar 4.25 Grafik hubungan k terhadap silhouette berdasarkan luas rumah...	48
Gambar 4.26 Daftar kecamatan pada cluster 1 berdasarkan luas rumah.....	50
Gambar 4.27 Daftar kecamatan pada cluster 2 berdasarkan luas rumah.....	50
Gambar 4.28 Daftar kecamatan pada cluster 3 berdasarkan luas rumah.....	51
Gambar 4.29 Daftar kecamatan pada cluster 4 berdasarkan luas rumah.....	52
Gambar 4.30 Daftar kecamatan pada cluster 5 berdasarkan luas rumah.....	52
Gambar 4.31 Grafik hubungan k terhadap SSE berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak dan diabetes keluarga.....	53

Gambar 4.32 Grafik hubungan k terhadap silhouette berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak dan diabetes keluarga.....	53
Gambar 4.33 Daftar kecamatan pada cluster 1 berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak, dan diabetes keluarga	55
Gambar 4.34 Daftar kecamatan pada cluster 2 berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak, dan diabetes keluarga	56
Gambar 4.35 Daftar kecamatan pada cluster 3 berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak, dan diabetes keluarga	57
Gambar 4.36 Daftar kecamatan pada cluster 4 berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak, dan diabetes keluarga	57
Gambar 4.37 Tampilan Halaman Beranda	71
Gambar 4.38 Tampilan Halaman Clustering	73
Gambar 4.39 Tampilan Peta di Halaman Clustering	74
Gambar 4.40 Tampilan Daftar Cluster di Halaman Clustering	74
Gambar 4.41 Tampilan peta saat salah satu cluster diklik	75
Gambar 4.42 Tampilan informasi detail cluster	75

DAFTAR TABEL

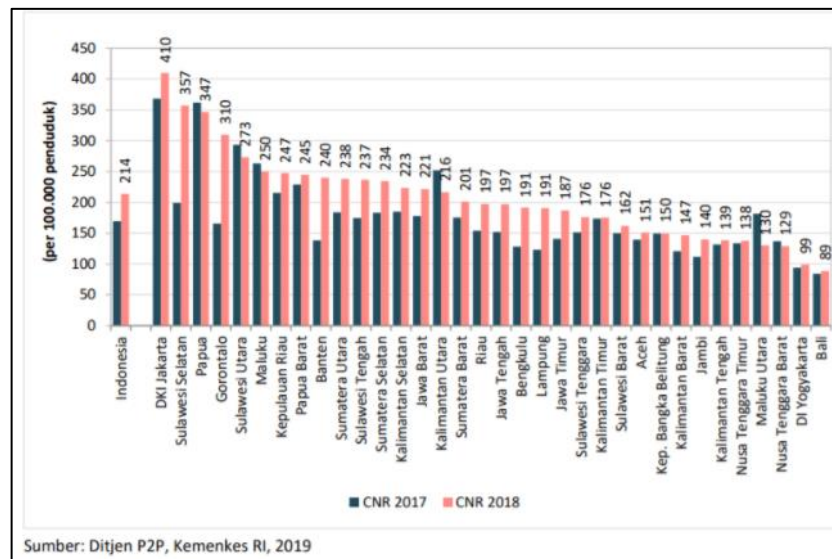
Tabel 4.1 Nilai k, SSE dan Silhouette berdasarkan jumlah tb dan usia	32
Tabel 4.2 Nilai statistik cluster berdasarkan jumlah tb dan usia	37
Tabel 4.3 Nilai k, SSE dan Silhouette berdasarkan status gizi anak	39
Tabel 4.4 Nilai statistik cluster berdasarkan status gizi	40
Tabel 4.5 Nilai k, SSE dan Silhouette berdasarkan pendapatan orang tua.....	43
Tabel 4.6 Nilai statistik cluster berdasarkan pendapatan orang tua	44
Tabel 4.7 Nilai k, SSE dan Silhouette berdasarkan luas rumah	49
Tabel 4.8 Nilai statistik cluster berdasarkan luas rumah.....	49
Tabel 4.9 Nilai k, SSE dan Silhouette berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak dan diabetes keluarga	54
Tabel 4.10 Nilai statistik cluster berdasarkan persentase telah BCG, ASI eksklusif, TB serumah, diabetes anak, dan diabetes keluarga.....	54
Tabel 4.11 Hasil clustering berdasarkan jumlah TB dan usia	58
Tabel 4.12 Hasil clustering berdasarkan status gizi	59
Tabel 4.13 Hasil clustering berdasarkan pendapatan orang tua	59
Tabel 4.14 Hasil clustering berdasarkan luas rumah.....	60
Tabel 4.15 Hasil clustering berdasarkan persentase anak telah BCG, ASI eksklusif, TB serumah, diabetes anak dan diabetes keluarga.....	61
Tabel 4.16 Interpretasi hasil clustering untuk setiap kecamatan di Makassar dan Gowa	64
Tabel 4.17 Data responden SUS.....	76
Tabel 4.18 Hasil Perhitungan skor SUS	77

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis (TB) adalah penyakit yang disebabkan oleh infeksi bakteri *Mycobacterium tuberculosis*. Sebagian besar kuman TB menyerang paru, tetapi dapat juga mengenai organ tubuh lainnya. Pada tahun 2017, Indonesia berada di peringkat ketiga setelah India dan China dengan jumlah penderita TB sebanyak 842 ribu kasus berdasarkan data dari *World Health Organization* (WHO).



Gambar 1.1 Angka Notifikasi Semua Kasus Tuberkulosis per 100.000 Penduduk Menurut Provinsi Tahun 2018

Gambar 1.1 menunjukkan bahwa Provinsi Sulawesi Selatan berada pada urutan kedua setelah DKI Jakarta berdasarkan angka notifikasi semua kasus tuberkulosis per 100.000 penduduk menurut provinsi tahun 2018. Tuberkulosis dapat menyerang segala kalangan usia, baik anak-anak, dewasa, maupun lansia. Masalah yang dihadapi pada tuberkulosis anak adalah kesulitan diagnosis, pengobatan, dan belum

optimalnya program pencegahan (Ketua UKK Respirologi IDAI, 2016). Mendeteksi tuberkulosis pada anak lebih sulit daripada orang dewasa karena pada kasus anak jarang menunjukkan gejala TB, misalnya batuk atau masalah lain pada saluran pernapasan. Tuberkulosis anak mempunyai permasalahan khusus yang berbeda dengan TB paru dewasa.

Salah satu permasalahan TB anak di Indonesia adalah penegakan diagnosis. Sejak tahun 2005 sistem skoring TB anak disosialisasikan dan direkomendasikan sebagai pendekatan diagnosis. Untuk mendiagnosis, diperlukan anamnesis yang teliti meliputi adanya kontak dengan TB dewasa aktif, gejala klinis yang sugestif termasuk evaluasi kurva pertumbuhan. Pemeriksaan fisis yang cermat harus dilakukan termasuk memeriksa tanda klinis yang spesifik sesuai TB. Pemeriksaan penunjang yang utama adalah uji kulit tuberkulin dan foto rontgen, namun saat ini pemeriksaan bakteriologis sangat penting diupayakan pada anak. Dengan menganalisis hasil pemeriksaan yang teliti dapat dihindari *overdiagnosis* atau *underdiagnosis* TB anak. Permasalahannya, tidak semua fasilitas pelayanan kesehatan (fasyankes) di Indonesia mempunyai fasilitas uji tuberkulin dan pemeriksaan foto toraks yang merupakan 2 parameter yang ada di sistem skoring. Akibatnya, di fasyankes dengan akses dan fasilitas terbatas banyak dijumpai *underdiagnosis* TB anak. Permasalahan lain dalam program penanggulangan TB anak adalah semakin meningkatnya jumlah kasus TB resistan obat pada dewasa, yang bisa merupakan sumber penularan bagi anak (Kementerian Kesehatan Republik Indonesia, 2016).

Oleh sebab itu, dengan banyaknya kasus TB di Indonesia terutama di Provinsi Sulawesi Selatan dan permasalahan khusus yang dihadapi oleh tuberkulosis anak, maka penelitian ini dilakukan untuk membantu dalam melakukan analisis terhadap penyakit tuberkulosis pada anak menggunakan *data mining*, sehingga diharapkan dapat memberikan informasi sebagai dasar untuk melakukan tindakan yang diperlukan dalam pengendalian pencegahan kasus tuberkulosis pada anak.

1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian ini adalah sebagai berikut:

- a) Bagaimana mengelompokkan data pasien tuberkulosis anak per kecamatan dengan *k-means clustering* dalam *data mining*?
- b) Bagaimana memvisualisasikan hasil *clustering data mining* ke dalam WebGIS?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

- a) Untuk menerapkan teknologi *data mining* dengan teknik *clustering* untuk mengelompokkan data pasien tuberkulosis anak per kecamatan berdasarkan kesamaan/kedekatan kasusnya.
- b) Untuk membuat sistem WebGIS yang akan memvisualisasikan hasil *clustering* ke bentuk peta digital.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

- a) Memberikan dorongan penggunaan teknologi *data mining* dengan teknik *clustering* untuk mendapatkan informasi pengelompokan data tuberkulosis pada anak per kecamatan.
- b) Memberikan informasi sebagai dasar pengambilan keputusan atau tindakan untuk pengendalian dan pencegahan tuberkulosis pada anak di masa yang akan datang.

1.5 Batasan Masalah Penelitian

- a) Pengambilan data dilakukan di Klinik Pediatrica Husada, yang berupa kuesioner yang dibagikan ke pasien tuberkulosis pada anak, yang kemudian akan dibentuk dalam format *excel* (.csv).
- b) Data yang dianalisis dalam penelitian ini ada data pasien tuberkulosis pada anak di Klinik Pediatrica Husada dari tahun 2018 hingga tahun 2022.

1.6 Sistematika Penulisan

Berikut adalah gambaran singkat mengenai isi tulisan secara keseluruhan:

BAB I PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas landasan teori yang digunakan untuk menganalisis masalah yang akan diteliti serta hal-hal lain yang berhubungan dengan variabel-variabel data yang akan digunakan, seperti *data mining*, *clustering*, algoritma *k-means*, serta web GIS.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi mengenai tahapan penelitian, instrumen penelitian, pengumpulan data, penerapan algoritma, teknik pengolahan data, serta visualisasi hasil *clustering* pada web GIS.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang sistem yang telah dibangun serta pembahasan dari hasil penelitian.

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.