



**PERBANDINGAN METODE K-MEANS DAN *COMPLETE LINKAGE*  
UNTUK PREDIKSI PENDERITA PENYAKIT TIROID MENGGUNAKAN  
KLASIFIKASI *XGBOOST***

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh**

**NANDHA AULIA EKAWATI**

**NIM 1711016120013**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JUNI 2023**



**PERBANDINGAN METODE K-MEANS DAN COMPLETE LINKAGE  
UNTUK PREDIKSI PENDERITA PENYAKIT TIROID MENGGUNAKAN  
KLASIFIKASI XGBOOST**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer**

**Oleh**

**NANDHA AULIA EKAWATI**

**NIM 1711016120013**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2023**

**SKRIPSI**

**PERBANDINGAN METODE K-MEANS DAN COMPLETE LINKAGE  
UNTUK PREDIKSI PENDERITA PENYAKIT TIROID MENGGUNAKAN  
KLASIFIKASI XGBOOST**

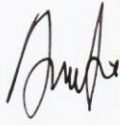
Oleh :

**NANDHA AULIA EKAWATI**

**NIM 1711016120013**

Telah dipertahankan di depan Dosen Penguji pada tanggal 19 Juni 2023. Susunan  
Dosen Penguji :

Pembimbing Utama



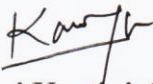
**Turiando Hamonangan Saragih,**  
**S.Kom., M.Kom**  
**NIP. 199308242019031012**

Dosen Penguji I



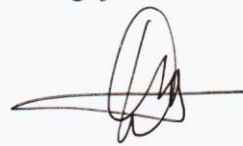
**Irwan Budiman, S.T., M.Kom.**  
**NIP. 197703252008121001**

Pembimbing Pendamping



**Dwi Kartini, S.Kom., M.Kom.**  
**NIP. 198704212012122003**

Dosen Penguji II



**Rahmat Ramadhani, S.Kom, M.Cs**  
**NIP. 19920330201901110001**



Banjarbaru, 19 Juni 2023  
Kelas Program Studi Ilmu Komputer,

**Irwan Budiman, S.T., M.Kom.**  
**NIP. 197703252008121001**

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, Juni 2023

Yang Menyatakan,



Nandha Aulia Ekawati

NIM 1711016120013

## ABSTRAK

### **PERBANDINGAN METODE K-MEANS DAN COMPLETE LINKAGE UNTUK PREDIKSI PENDERITA PENYAKIT TIROID MENGGUNAKAN KLASIFIKASI XGBOOST**

(Oleh: Nandha Aulia Ekawati; Pembimbing: Triando Hamonangan Saragih S.Kom., M.Kom dan Dwi Kartini, S.Kom, M.Kom.; 2023; 87 halaman)

Kelenjar tiroid bisa disebut sebagai organ paling penting dalam tubuh, karena hormon tiroid berfungsi untuk mengendalikan metabolisme dalam tubuh. Sehingga jika kelenjar tiroid mengalami kelainan maka metabolisme dalam tubuh juga akan terganggu. Gangguan tiroid dapat dibedakan menjadi 3 jenis menurut fungsinya, yaitu *hypothyroid*, *hyperthyroid*, dan *euthyroid*. Untuk membantu memprediksi penyakit dari kelainan tiroid, pada penelitian ini digunakan metode *clustering k-means* dan *complete linkage* untuk mengelompokkan data menjadi 3 kelompok dan menggunakan metode klasifikasi *XGBoost* untuk melihat tingkat akurasi dari pengelompokkan data metode-metode *clustering*. Pada penelitian ini melakukan perbandingan tingkat akurasi dari klasifikasi *XGBoost* dengan menggunakan metode *clustering k-means* dan klasifikasi *XGBoost* dengan menggunakan metode *clustering complete linkage*. Berdasarkan hasil penelitian, didapatkan bahwa hasil akurasi dari klasifikasi *XGBoost* dengan metode *clustering k-means* menghasilkan kinerja akurasi yang lebih baik, yaitu sebesar 99,957%. Sedangkan hasil akurasi dari klasifikasi *XGBoost* menggunakan metode *clustering complete linkage* sebesar 99,67%.

**Kata kunci :** Penyakit Tiroid, *Clustering*, *K-Means*, *Complete Linkage*, *Klasifikasi*, *XGBoost*

## **ABSTARCT**

### **COMPARISON OF K-MEANS AND COMPLETE LINKAGE CLUSTERING METHODS TO PREDICT THYROID DISEASE PATIENTS WITH XGBOOST CLASSIFICATION**

*(By: Nandha Aulia Ekawati; Advisors: Triando Hamonangan Saragih S.Kom., M.Kom dan Dwi Kartini, S.Kom, M.Kom.; 2023; 87 pages)*

*The thyroid gland can be called the most important organ in the body, because thyroid hormones function to control metabolism in the body. So if the thyroid gland is abnormal, the metabolism in the body will also be disturbed. Thyroid disorders can be divided into 3 types according to their function, namely hypothyroid, hyperthyroid, and euthyroid. To help predict diseases from thyroid disorders, this study used k-means clustering and complete linkage methods to group data into 3 groups and used the XGBoost classification method to see the accuracy level of data clustering methods. This study compares the accuracy of the XGBoost classification using the k-means clustering method and the XGBoost classification using the complete linkage clustering method. Based on the results of the study, it is found that the accuracy of the XGBoost classification with the k-means clustering method produces better accuracy performance, which is 99.957%. While the accuracy of the XGBoost classification using the complete linkage clustering method is 99.67%.*

**Keywords:** *Thyroid Disease, Clustering, K-Means, Complete Linkage, Classification, XGBoost*

## PRAKATA

Assalamualaikum warahmatullahi wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Metode *K-Means* dan *Complete Linkage* untuk Prediksi Penderita Penyakit Tiroid Menggunakan Klasifikasi *XGBoost*”, untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah*.

Tidak lupa penulis mengucapkan terima kasih kepada pihak-pihak terkait yang sudah membantu dalam penyelesaian skripsi ini, diantaranya:

1. Orang tua tercinta untuk ibu dan almarhum bapak yang dari awal proses perkuliahan selalu memberikan bantuan, semangat, doa dan dukungan hingga sampai pada tahap proses penyelesaian skripsi ini.
2. Bapak Triando Hamonangan Saragih, S. Kom., M. Kom. selaku dosen pembimbing utama dan Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang telah memberikan waktu dan bimbingannya kepada penulis dalam pengerjaan skripsi.
3. Sahabat magurl Tari, Dhea, Sally, Tia, Atul, Vina, Huda yang saling memberikan semangat dan dukungan untuk kelancaran penyelesaian skripsi ini yang telah berjuang bersama dari awal perkuliahan.
4. Teman yang selalu mau menjadi teman diskusi serta membantu menjawab setiap pertanyaan saya dan membantu saya mencari solusi akan permasalahan yang saya dapatkan selama mengerjakan skripsi.
5. Teman – teman Ilmu Komputer 2017 yang masuk jurusan ini bersama saya, terimakasih atas tangis dan tawanya.
6. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2023

Nandha Aulia Ekawati



## DAFTAR ISI

SKRIPSI.....	ii
PERNYATAAN.....	iii
ABSTRAK .....	iv
<i>ABSTARCT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	2
1.3    Batasan Masalah.....	3
1.4    Tujuan.....	3
1.5    Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1.    Kajian Terdahulu .....	4
2.2.    Landasan Teori .....	6
2.2.1    Penyakit Tiroid.....	6
2.2.2    Data Mining .....	7
2.2.3    Min-Max Normalization .....	9
2.2.4    Non-hierarchical dan Hierarchical Clustering .....	10
2.2.5 <i>K-Means</i> .....	11
2.2.6 <i>Complete Linkaage</i> .....	12
2.2.7 <i>Silhouette</i> .....	13
2.2.8 <i>Boosting</i> .....	14

2.2.9	<i>Extreme Gradient Boosting</i> .....	15
2.2.10	<i>Cross Validation</i> .....	16
2.2.11	Evaluasi.....	17
BAB III METODE PENELITIAN.....		20
3.1	Alat Penelitian .....	20
3.2	Bahan Penelitian.....	20
3.3	Variabel Penelitian .....	22
3.4	Prosedur Penelitian.....	22
BAB IV HASIL DAN PEMBAHASAN .....		25
4.1	Hasil.....	25
4.1.1	Pengumpulan Data .....	25
4.1.2	<i>Preprocessing Data</i> .....	26
4.1.3	<i>Clustering Data</i> .....	28
4.1.4	Evaluasi <i>Clustering</i> .....	37
4.1.5	Pembagian Data .....	38
4.1.6	Klasifikasi .....	39
4.1.7	<i>Extreme Gradient Boosting</i> dengan <i>Clustering K-means</i> .....	44
4.1.8	<i>Extreme Gradient Boosting</i> dengan <i>Clustering Complete Linkage</i> .....	54
4.1.9	Evaluasi.....	63
4.2	Pembahasan .....	82
BAB V PENUTUP.....		87
5.1	Kesimpulan.....	87
5.2	Saran.....	87
DAFTAR PUSTAKA .....		88
LAMPIRAN		

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1 Keaslian Penelitian.....	5
Tabel 2 Perancangan Penelitian .....	6
Tabel 3 <i>Confusion matrix</i> .....	17
Tabel 4 Kategori Klasifikasi berdasarkan Nilai AUC.....	18
Tabel 5 Atribut kategorikal dataset penyakit tiroid.....	20
Tabel 6 Atribut numerikal data set penyakit tiroid .....	21
Tabel 7 Contoh dataset Penyakit Tiroid.....	21
Tabel 8 Nama atribut dan keterangannya.....	25
Tabel 9 Contoh dataset Penyakit Tiroid.....	26
Tabel 10 Contoh Perhitungan Normalisasi min-max .....	27
Tabel 11 Contoh data ternormalisasi.....	27
Tabel 12 Penentuan <i>centroid</i> awal .....	29
Tabel 13 Perhitungan <i>Euclidean distance</i> untuk menentukan jarak terdekat .....	30
Tabel 14 Penetapan jarak <i>cluster</i> ke titik data pada iterasi 0 .....	30
Tabel 15 Perhitungan <i>centroid</i> baru.....	31
Tabel 16 <i>Centroid</i> baru pada iterasi 1 .....	31
Tabel 17 <i>Centroid</i> terakhir pada iterasi 6.....	32
Tabel 18 Penetapan jarak <i>cluster</i> ke titik data pada iterasi 6 .....	33
Tabel 19 Hasil <i>labeling clustering K-means</i> .....	34
Tabel 20 Contoh perhitungan jarak objek.....	35
Tabel 21 Hasil matrix jarak.....	35
Tabel 22 Contoh perhitungan antar kelompok.....	36
Tabel 23 Hasil matrix jarak setelah perhitungan pertama.....	36
Tabel 24 Hasil <i>labeling clustering complete linkage</i> .....	37
Tabel 25 Pembagian <i>K-fold Cross Validation</i> .....	38
Tabel 26 Contoh data .....	39
Tabel 27 Perhitungan residual.....	40
Tabel 28 Hasil <i>split</i> data.....	40

Tabel 29 Hasil perhitungan <i>similarity</i> dan <i>gain</i> .....	41
Tabel 30 Hasil perhitungan <i>similarity</i> dan <i>gain</i> iterasi 2 .....	42
Tabel 31 Hasil perhitungan <i>similarity</i> dan <i>gain</i> iterasi 3 .....	43
Tabel 32 Hasil akurasi dan AUC <i>XGBoost</i> dari 1-10 <i>fold cross validation</i> dengan metode <i>clustering K-means</i> dengan parameter <i>default</i> .....	44
Tabel 33 Nilai akurasi metode <i>clustering K-means</i> untuk parameter <i>Max Depth</i> .....	46
Tabel 34 Nilai AUC metode <i>clustering K-means</i> untuk parameter <i>Max Depth</i> .....	46
Tabel 35 Nilai akurasi metode <i>clustering K-means</i> untuk parameter <i>Gamma</i> .....	47
Tabel 36 Nilai AUC metode <i>clustering K-means</i> untuk parameter <i>Gamma</i> .....	48
Tabel 37 Nilai akurasi metode <i>clustering K-means</i> untuk parameter <i>Learning Rate</i> .....	49
Tabel 38 Nilai AUC metode <i>clustering K-means</i> untuk parameter <i>Learning Rate</i> .....	49
Tabel 39 Nilai akurasi metode <i>clustering K-means</i> untuk parameter <i>N estimator</i> .....	50
Tabel 40 Nilai AUC metode <i>clustering K-means</i> untuk parameter <i>N estimator</i> .....	51
Tabel 41 Hasil akurasi dan AUC <i>XGBoost</i> dari 1-10 <i>fold cross validation</i> dengan metode <i>clustering K-means</i> dengan parameter terbaik .....	52
Tabel 42 Hasil akurasi dan AUC <i>XGBoost</i> dari 1-10 <i>fold cross validation</i> dengan metode <i>clustering complete linkage</i> .....	54
Tabel 43 Nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Max Depth</i> .....	56
Tabel 44 Nilai AUC metode <i>clustering complete linkage</i> untuk parameter <i>Max Depth</i> .....	56
Tabel 45 Nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Gamma</i> .....	57
Tabel 46 Nilai AUC metode <i>clustering complete linkage</i> untuk parameter <i>Gamma</i> .....	58
Tabel 47 Nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Learning Rate</i> .....	59
Tabel 48 Nilai AUC metode <i>clustering Complete linkage</i> untuk parameter <i>Learning Rate</i> .....	59
Tabel 49 Nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>N estimator</i> .....	60

Tabel 50 Nilai AUC metode <i>clustering complete linkage</i> untuk parameter <i>N estimator</i> .....	61
Tabel 51 Hasil akurasi dan AUC <i>XGBoost</i> dari 1-10 <i>fold cross validation</i> dengan metode <i>clustering complete linkage</i> dengan parameter terbaik .....	62
Tabel 52 Perhitungan akurasi klasifikasi <i>XGBoost clustering K-means</i> parameter <i>default</i> .....	66
Tabel 53 Perhitungan akurasi klasifikasi <i>XGBoost clustering K-means</i> parameter terbaik.....	70
Tabel 54 Perhitungan akurasi klasifikasi <i>XGBoost clustering complete linkage</i> parameter <i>default</i> .....	75
Tabel 55 Perhitungan akurasi klasifikasi <i>XGBoost clustering complete linkage</i> parameter terbaik.....	79
Tabel 56 Perbandingan hasil AUC sebelum dan sesudah uji parameter.....	85
Tabel 57 Perbandingan hasil akurasi sebelum dan sesudah uji parameter.....	86

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1 Tahapan Data Mining.....	8
Gambar 2 Perbandingan <i>Hierarchical clustering</i> dan <i>Partitional clustering</i> .....	10
Gambar 3 <i>Complete Linkage</i> antara dua obyek .....	12
Gambar 4 Model <i>boosting</i> .....	14
Gambar 5 Alur kerja dari Cross validation .....	17
Gambar 6 Alur Penelitian.....	23
Gambar 7 Hasil plotting K-means dengan <i>cluster</i> berjumlah 3 .....	34
Gambar 8 Dendrogram <i>complete linkage</i> dengan <i>cluster</i> berjumlah 3 .....	37
Gambar 9 Membangun pohon dengan pembagian pohon .....	41
Gambar 10 Pohon dengan <i>gain</i> optimal iterasi 1 .....	42
Gambar 11 Pohon dengan <i>gain</i> optimal iterasi 2 .....	43
Gambar 12 Pohon dengan <i>gain</i> optimal iterasi 3 .....	43
Gambar 13 Pohon Keputusan <i>XGBoost</i> dengan metode clustering K-means (a) ke-1, (b) ke-2, dan (c) terakhir dengan parameter <i>default</i> .....	45
Gambar 14 Grafik nilai akurasi metode <i>clustering</i> K-means untuk parameter <i>Max Depth</i> .....	47
Gambar 15 Grafik nilai akurasi metode <i>clustering</i> K-means untuk parameter <i>Gamma</i> .....	48
Gambar 16 Grafik nilai akurasi metode <i>clustering</i> K-means untuk parameter <i>Learning Rate</i> .....	50
Gambar 17 Grafik nilai akurasi metode <i>clustering</i> K-means untuk parameter <i>N estimator</i> .....	51
Gambar 18 Pohon Keputusan <i>XGBoost</i> dengan metode clustering K-means (a) ke-1, (b) ke-2, dan (c) terakhir setelah uji parameter .....	53
Gambar 19 Pohon Keputusan <i>XGBoost</i> dengan metode clustering <i>complete linkage</i> (a) ke-1, (b) ke-2, dan (c) terakhir dengan parameter default.....	55
Gambar 20 Grafik nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Max Depth</i> .....	57

Gambar 21 Grafik nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Gamma</i> .....	58
Gambar 22 Grafik nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>Learning Rate</i> .....	60
Gambar 23 Grafik nilai akurasi metode <i>clustering complete linkage</i> untuk parameter <i>N estimator</i> .....	61
Gambar 24 Pohon Keputusan <i>XGBoost</i> dengan metode <i>clustering complete linkage</i> (a) ke-1, (b) ke-2, dan (c) terakhir setelah uji parameter .....	63
Gambar 25 <i>Confusion matrix 10-fold cross validation XGBoost</i> dengan metode <i>clustering K-means</i> dengan parameter <i>default</i> .....	65
Gambar 26 Kurva AUC <i>10-fold cross validation XGBoost</i> dengan metode <i>clustering K-means</i> dengan parameter <i>default</i> .....	68
Gambar 27 <i>Confusion matrix 10-fold cross validation XGBoost</i> dengan metode <i>clustering K-means</i> setelah uji parameter .....	70
Gambar 28 Kurva AUC <i>10-fold cross validation XGBoost</i> dengan metode <i>clustering K-means</i> setelah uji parameter .....	72
Gambar 29 <i>Confusion matrix 10-fold cross validation XGBoost</i> dengan metode <i>clustering complete linkage</i> .....	74
Gambar 30 Kurva AUC <i>10-fold cross validation XGBoost</i> dengan metode <i>clustering complete linkage</i> dengan parameter <i>default</i> .....	77
Gambar 31 <i>Confusion matrix 10-fold cross validation XGBoost</i> dengan metode <i>clustering Complete linkage</i> setelah uji parameter .....	79
Gambar 32 Kurva AUC <i>10-fold cross validation XGBoost</i> dengan metode <i>clustering Complete linkage</i> setelah uji parameter .....	81
Gambar 33 Perbandingan jumlah data per kelas hasil <i>clustering</i> .....	83
Gambar 34 Perbandingan hasil evaluasi <i>Confusion Matrix</i> menggunakan parameter <i>default</i> .....	84
Gambar 35 Perbandingan hasil evaluasi <i>Confusion Matrix</i> menggunakan parameter terbaik .....	85

## DAFTAR LAMPIRAN

### Lampiran

Lampiran 1 Data Asli

Lampiran 2 *Source Code Clustering K-means*

Lampiran 3 *Source Code Clustering Complete Linkage*

Lampiran 4 *Source Code Klasifikasi XGBoost*

Lampiran 5 Riwayat Hidup