

# **SKRIPSI**

## **IMPLEMENTASI DATA MINING UNTUK *CLUSTERING* INDEKS PEMBANGUNAN MANUSIA (IPM) BERDASARKAN KABUPATEN/KOTA DI PROVINSI NTT MENGGUNAKAN METODE *K-MEANS* DAN *WARD LINKAGE***



Oleh :

**MARIA RELY SUNENSI**

NIM: 19083000115

**PROGRAM S1 SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI**

**UNIVERSITAS MERDEKA**

**MALANG**

**2023**

**IMPLEMENTASI DATA MINING UNTUK *CLUSTERING*  
INDEKS PEMBANGUNAN MANUSIA (IPM) BERDASARKAN  
KABUPATEN/KOTA DI PROVINSI NTT MENGGUNAKAN  
METODE *K-MEANS* DAN *WARD LINKAGE***



**SKRIPSI**

Diajukan Sebagai Salah Satu Syarat  
Memperoleh Gelar Sarjana Sistem Informasi  
Pada Program S1 Sistem Informasi  
Universitas Merdeka Malang

Oleh :

**MARIA RELY SUNENSI**

NIM: 19083000115

**PROGRAM S1 SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS MERDEKA  
MALANG**

**2023**

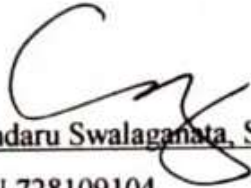
## LEMBAR PERSETUJUAN

Nama : Maria Rely Sunensi  
NIM : 19083000115  
Universitas : Universitas Merdeka Malang  
Fakultas : Teknologi Informasi  
Program Studi : SI Sistem Informasi  
Judul : Implementasi Data Mining Untuk *Clustering*  
Indeks Pembangunan Manusia (IPM) Berdasarkan  
Kabupaten/Kota Di Provinsi NTT Menggunakan  
Metode *K-Means* Dan *Ward Linkage*


Malang, 24 Juli 2023

Disetujui Dan Diterima

Ketua Program Studi SI Sistem Informasi


  
Galandaru Swalaganata, S.Si., M.Si  
NIDN.728109104

Dosen Pembimbing

  
Drs. Anis Zubair, M.Kom.  
NIDN.701116703

Dekan

Fakultas Teknologi Informasi

  
Dr. Mardiana Andrawati, SE., MSi.  
NIDN.716037601

**IMPLEMENTASI DATA MINING UNTUK *CLUSTERING*  
INDEKS PEMBANGUNAN MANUSIA (IPM) BERDASARKAN  
KABUPATEN/KOTA DI PROVINSI NTT MENGGUNAKAN  
METODE *K-MEANS* DAN *WARD LINKAGE***

Dipersiapkan dan disusun oleh

Maria Rely Sunensi

19083000115

Telah dipertahankan di depan Dewan Penguji

Pada Tanggal 24 Juli 2023

Susunan Dewan Penguji

Ketua Penguji



Luthfi Indana, S.Pd., M.Pd.  
NIDN.711059203

Sekretaris Penguji



Drs. Anis Zubair, M.Kom.  
NIDN.701116703

Anggota Penguji



Rahmatina Hidayati, S.Kom., M.T.  
NIDN.720028902

Skripsi Ini Telah Diterima Sebagai Salah Satu Persyaratan untuk  
Memperoleh Gelar Sarjana Sistem Informasi

Malang 24 Juli 2023

Dekan Fakultas Teknologi Informasi

Universitas Merdeka Malang



Desa Anindyaning Andarwati, SE., M.Si.  
NIDN.716037601

## SURAT PERNYATAAN

Yang bertanda tangan dibawah ini

Nama : Maria Rely Sunensi  
NIM : 19083000115  
Program Studi : S1 Sistem Informasi  
Bidang Kajian Skripsi : Analisis Data  
Judul Tugas Akhir : Implementasi Data Mining Untuk *Clustering*  
Indeks Pembangunan Manusia (IPM) Berdasarkan  
Kabupaten/Kota Di Provinsi NTT Menggunakan  
Metode *K-Means* Dan *Ward Linkage*  
Lokasi Penelitian : Badan Pusat Statistik Nusa Tenggara Timur  
Alamat Rumah Asal : Jln. Kelapa Sawit No.18  
No. Telp/Email : 081230674028  
relysunensi25@gmail.com

Dengan ini menyatakan bahwa saya benar-benar melakukan penelitian dan penulisan skripsi tersebut diatas benar-benar karya saya dan tidak melakukan plagiasi. kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam dengan disebutkan nama dan dicantumkan dalam daftar pustaka.

Demikian surat pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan tidak kebenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan aturan yang berlaku.

Malang, 24 Juli 2023

Peneliti,



Maria Rely Sunensi

## **MOTO**

Tidak Ada Akhir Yang Gagal  
Untuk Orang Yang Hidupnya Mengandalkan Tuhan  
Semua Akan Indah Pada Waktunya

## **LEMBAR PERSEMBAHAN**

Skripsi ini penulis persembahkan untuk kedua orang tua, kakak, adik dan keluarga yang sudah memberikan dukungan penuh bagi penulis dari masih kecil hingga sekarang. Yang sudah dengan susah payah dalam pengorbanan, doa, usaha hingga cinta dan kasih sayang yang telah diberikan. Serta tak lupa untuk semua sahabat dekat penulis yang ikut membantu dalam memberi dukungan kepada penulis. Terima kasih banyak penulis ucapkan kepada seluruh orang-orang yang telah memberikan dukungan penuh.

## KATA PENGANTAR

Puji dan Syukur penulis haturkan kehadiran Tuhan Yang Maha Esa, karena berkat Karunia-Nya, tugas akhir ini dapat diselesaikan. Pada kesempatan yang baik ini dan dengan penuh rasa syukur yang mendalam, perkenankan penulis mengucapkan terimakasih yang sebesar-besarnya dan apresiasi yang setinggi-tingginya kepada yang terhormat:

1. Ibu Dr. Mardiana Andarwati, SE., M.Si selaku Ketua Dekan Program Studi Sistem Informasi yang telah memberikan dedikasi selama 4 tahun untuk tercapainya pembelajaran maksimal bagi program studi yang dipimpin.
2. Bapak Galandaru Swalaganata, S.Si., M.Si selaku Ketua Program Studi S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
3. Bapak Drs. Anis Zubair, M.Kom selaku dosen pembimbing saya yang telah memberikan arahan, bimbingan, serta motivasi kepada penulis dan terima kasih atas kesabaran beliau kepada penyusun selama proses pengerjaan skripsi ini.
4. Kedua Orangtua, kak Jefri, kak Meri, kak Denti, adik Hilda juga Chantika dan Geonatan yang senantiasa menjadi alarm pengingat untuk makan dan tidur juga sebagai penyemangat dalam pengerjaan skripsi ini.
5. Kiyut Family yang telah membantu dan memberi semangat selama proses pengerjaan skripsi.
6. Teman kelas dan saudara di tanah rantauan yang sudah berjuang bersama dengan caranya masing-masing membantu penulis selama menyusun skripsi.

## DAFTAR ISI

HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	ii
LEMBAR PERSETUJUAN.....	iii
LEMBAR PENGESAHAN .....	iv
SURAT PERNYATAAN .....	v
MOTO.....	vi
LEMBAR PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
DAFTAR LAMPIRAN.....	xiv
ABSTRAK .....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.4.1 Manfaat bagi Peneliti .....	3
1.4.2 Manfaat bagi Lembaga.....	3
1.5 Batasan Penelitian.....	3
BAB II KAJIAN PUSTAKA.....	4
2.1 Landasan Teori.....	4
2.1.1 IPM (Indeks Pembangunan Manusia).....	4
2.1.2 <i>Data Mining</i> .....	5
2.1.3 <i>Clustering</i> .....	7
2.1.4 <i>K-Means</i> .....	7
2.1.5 <i>Ward Linkage</i> .....	8
2.1.6 <i>Orange</i> .....	8

2.2	Penelitian Terdahulu .....	10
BAB III METODE PENELITIAN .....		12
3.1	Desain Penelitian .....	12
3.2	Lokasi Penelitian.....	13
3.3	Populasi dan Sampel.....	13
3.4	Teknik Pengumpulan Data.....	13
3.4.1	BPS (Badan Pusat Statistik).....	13
3.4.2	Studi Pustaka.....	13
3.5	Teknik Analisis Data.....	14
BAB IV HASIL DAN PEMBAHASAN .....		16
4.1	Hasil Penelitian .....	16
4.2	Pembahasan.....	17
4.2.1	<i>Clustering</i> Menggunakan Metode <i>K-Means</i> .....	18
4.2.2	<i>Clustering</i> Menggunakan Metode <i>Ward Linkage</i> .....	42
4.2.3	Analisis hasil .....	50
BAB V PENUTUP.....		53
5.1	Kesimpulan .....	53
5.2	Saran .....	53
DAFTAR PUSTAKA.....		54
LAMPIRAN.....		56

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	10
Tabel 3. 1 Alur Metode <i>K-Means</i> dan <i>Ward Linkage</i> .....	14
Tabel 4. 1 Penentuan Kriteria Korelasi .....	17
Tabel 4. 2 Uji Korelasi antara IPM dan Ril per Kapita.....	18

## DAFTAR GAMBAR

Gambar 2. 1 Tampilan Awal pada Aplikasi <i>Orange</i> .....	9
Gambar 3. 1 Desain Penelitian.....	12
Gambar 4. 1 Data IPM dan Pengeluaran Ril per Kapita .....	16
Gambar 4. 2 Tampilan untuk Metode <i>K-Means</i> .....	18
Gambar 4. 3 Penentuan <i>Cluster</i> Tahun 2018 .....	19
Gambar 4. 4 Hasil <i>Clustering</i> Data <i>Table</i> Tahun 2018.....	20
Gambar 4. 5 Tampilan Peta Tahun 2018 .....	20
Gambar 4. 6 Penentuan <i>Cluster</i> Tanpa Data Kota Kupang .....	21
Gambar 4. 7 Hasil <i>Clustering</i> Tanpa Data Kota Kupang.....	22
Gambar 4. 8 Penentuan <i>Cluster</i> Tanpa Data Sabu Raijua .....	22
Gambar 4. 9 Hasil <i>Clustering</i> Tanpa Data Sabu Raijua.....	23
Gambar 4. 10 Tampilan Peta Tanpa <i>Outlier</i> Tahun 2018.....	24
Gambar 4. 11 Penentuan <i>Cluster</i> Tahun 2019 .....	24
Gambar 4. 12 Hasil <i>Clustering K-Means</i> Tahun 2019.....	25
Gambar 4. 13 Tampilan Peta Tahun 2019 .....	25
Gambar 4. 14 Penentuan <i>Cluster</i> Tanpa Kota Kupang .....	26
Gambar 4. 15 Hasil <i>Clustering</i> Tanpa Data Kota Kupang.....	27
Gambar 4. 16 Penentuan <i>Cluster</i> tanpa data Sabu Raijua.....	27
Gambar 4. 17 Hasil <i>Clustering</i> Tanpa Data Sabu Raijua.....	28
Gambar 4. 18 Tampilan Peta tanpa <i>Outlier</i> Tahun 2019 .....	29
Gambar 4. 19 Penentuan <i>Cluster</i> Tahun 2020 .....	29
Gambar 4. 20 Hasil <i>Clustering</i> Tahun 2020 .....	30
Gambar 4. 21 Tampilan Peta Tahun 2020 .....	31
Gambar 4. 22 Penentuan <i>Cluster</i> Tanpa Data Kota Kupang .....	31
Gambar 4. 23 Hasil <i>Clustering</i> Tanpa Kota Kupang .....	32
Gambar 4. 24 Tampilan Peta tanpa <i>Outlier</i> Tahun 2020 .....	33
Gambar 4. 25 Penentuan <i>Cluster</i> Tahun 2021 .....	33
Gambar 4. 26 Hasil <i>Clustering</i> Tahun 2021 .....	34
Gambar 4. 27 Tampilan Peta Tahun 2021 .....	35
Gambar 4. 28 Penentuan <i>Cluster</i> Tanpa Kota Kupang .....	35
Gambar 4. 29 Hasil <i>Clustering</i> Tanpa Data Kota Kupang.....	36
Gambar 4. 30 Tampilan Peta tanpa <i>Outlier</i> Tahun 2021 .....	36
Gambar 4. 31 Penentuan <i>Cluster</i> Tahun 2022 .....	37
Gambar 4. 32 Hasil <i>Clustering</i> Tahun 2022 .....	38
Gambar 4. 33 Tampilan Peta Tahun 2022 .....	38
Gambar 4. 34 Penentuan <i>Cluster</i> Tanpa Data Kota Kupang .....	39
Gambar 4. 35 Hasil <i>Clustering</i> Tanpa Data Kota Kupang.....	40
Gambar 4. 36 Penentuan <i>Cluster</i> Tanpa Data Sabu Raijua .....	40
Gambar 4. 37 Hasil <i>Clustering</i> Tanpa Data Sabu Raijua.....	41

Gambar 4. 38 Tampilan Peta tanpa <i>Outlier</i> tahun 2022.....	42
Gambar 4. 39 Tampilan Untuk Metode <i>Ward Linkage</i> .....	42
Gambar 4. 40 Hasil <i>Clustering</i> Metode <i>Ward Linkage</i> Tahun 2018 .....	43
Gambar 4. 41 Maps Tahun 2018 menggunakan <i>Ward Linkage</i> .....	44
Gambar 4. 42 Hasil <i>Clustering</i> Metode <i>Ward Linkage</i> Tahun 2019 .....	44
Gambar 4. 43 Maps Tahun 2019 Menggunakan <i>Ward Linkage</i> .....	45
Gambar 4. 44 Hasil <i>Clustering</i> Metode <i>Ward Linkage</i> Tahun 2020 .....	46
Gambar 4. 45 Maps Tahun 2020 Menggunakan <i>Ward Linkage</i> .....	46
Gambar 4. 46 Hasil <i>Clustering</i> Metode <i>Ward Linkage</i> Tahun 2021 .....	47
Gambar 4. 47 Maps Tahun 2021 Menggunakan <i>Ward Linkage</i> .....	48
Gambar 4. 48 Hasil <i>Clustering</i> Metode <i>Ward Linkage</i> Tahun 2022 .....	49
Gambar 4. 49 Maps Tahun 2022 Menggunakan <i>Ward Linkage</i> .....	49
Gambar 4. 50 Analisis hasil dari metode K-Means dan <i>Ward Linkage</i> .....	52

## DAFTAR LAMPIRAN

Lampiran 1. Halaman Riwayat Hidup .....	56
Lampiran 2. Hasil Plagiasi .....	57
Lampiran 3. Lembar Persetujuan Tugas Akhir .....	58

# IMPLEMENTASI DATA MINING UNTUK *CLUSTERING* INDEKS PEMBANGUNAN MANUSIA (IPM) BERDASARKAN KABUPATEN/KOTA DI PROVINSI NTT MENGGUNAKAN METODE *K-MEANS* DAN *WARD LINKAGE*

Maria Rely Sunensi

Program Studi Teknologi Informasi, Program  
Sistem Informasi Universitas Merdeka Malang

Email : [relsunensi25@gmail.com](mailto:relsunensi25@gmail.com)

## ABSTRAK

Pada tahun 2021 Indeks pembangunan manusia Nusa Tenggara Timur (NTT) meningkat dibandingkan sebelumnya. Di tahun 2021 IPM Provinsi Nusa Tenggara Timur sebesar 65,28 meningkat 0,14 persen (naik 0,09 poin) dibandingkan pada tahun 2020. Peningkatan IPM 2021 didukung oleh komponen kesehatan dan pendidikan, tetapi pada komponen pengeluaran per kapita yang disesuaikan menurun sebesar 0,58 persen dibanding tahun 2020. Metode yang digunakan dalam penelitian ini adalah K-Means dan Ward Linkage. Penelitian ini dilakukan untuk menghitung hubungan antara IPM dan Pengeluaran RII per Kapita juga untuk menghitung klasterisasi IPM berdasarkan Pengeluaran RII per Kapita di NTT menggunakan metode K-Means dan melakukan perbandingan dengan metode ward Linkage. Hasil perhitungan hubungan antara IPM dan Pengeluaran RII per menghasilkan keeratan hubungan yang sangat kuat yaitu Clustering K-Means dan pada tahun 2018 dan 2019 sama, C1 8 kabupaten (cluster rendah), C2 8 kabupaten (cluster sedang) dan C3 4 kabupaten (cluster tinggi). Di tahun 2020 dan 2021 sama, C1 8 kabupaten (cluster rendah), C2 12 kabupaten (cluster tinggi). Sedangkan pada tahun 2022 C1 9 kabupaten (cluster rendah), C2 8 kabupaten (cluster sedang), dan C3 3 kabupaten (cluster tinggi). Clustering Ward Linkage pada tahun 2018 C1 8 Kabupaten (cluster rendah), C2 9 kabupaten (cluster sedang) dan C3 3 kabupaten (cluster tinggi). Di tahun 2019 C1 9 kabupaten (cluster rendah), C2 8 kabupaten (cluster sedang) dan C3 3 kabupaten (cluster tinggi). Di tahun 2020 C1 9 kabupaten (cluster rendah), C2 12 kabupaten (cluster tinggi). Di tahun 2021 C1 10 kabupaten (cluster rendah), C2 11 kabupaten (cluster tinggi). Sedangkan di tahun 2022 C1 9 kabupaten (cluster rendah), C2 8 kabupaten (cluster sedang) dan C3 3 kabupaten (cluster tinggi).

Kata Kunci : IPM, *Clustering*, *K-Means*, *Ward Linkage*

# IMPLEMENTATION OF DATA MINING FOR HUMAN DEVELOPMENT INDEX (HDI) CLUSTERING BY DISTRICT/CITY IN NTT PROVINCE USING K-MEANS AND WARD LINKAGE METHODS

Maria Rely Sunensi

Program Studi Teknologi Informasi, Program

Sistem Informasi Universitas Merdeka Malang

Email : [relsunensi25@gmail.com](mailto:relsunensi25@gmail.com)

## ABSTRACT

In 2021, the East Nusa Tenggara (NTT) human development index increased compared to before. In 2021, the HDI of East Nusa Tenggara Province was 65.28, an increase of 0.14 percent (up 0.09 points) compared to 2020. The 2021 HDI increase was supported by the health and education components, but the adjusted per capita expenditure component decreased by 0.58 percent compared to 2020. The methods used in this research are K-Means and Ward Linkage. This research was conducted to calculate the relationship between HDI and real per capita expenditure as well as to calculate the clustering of HDI based on real per capita expenditure in NTT using the K-Means method and to make a comparison with the Ward Linkage method. The results of the calculation of the relationship between HDI and RII Expenditure per produces a very strong relationship, namely K-Means Clustering and in 2018 and 2019 the same, C1 8 districts (low cluster), C2 8 districts (medium cluster) and C3 4 districts (high cluster). In 2020 and 2021 the same, C1 8 districts (low cluster), C2 12 districts (high cluster). While in 2022 C1 9 districts (low cluster), C2 8 districts (medium cluster), and C3 3 districts (high cluster). Ward Linkage Clustering in 2018 C1 8 districts (low cluster), C2 9 districts (medium cluster) and C3 3 districts (high cluster). In 2019 C1 9 districts (low cluster), C2 8 districts (medium cluster) and C3 3 districts (high cluster). In 2020 C1 9 districts (low cluster), C2 12 districts (high cluster). In 2021 C1 10 districts (low cluster), C2 11 districts (high cluster). While in 2022 C1 9 districts (low cluster), C2 8 districts (medium cluster) and C3 3 districts (high cluster).

*Keyword : HDI, Clustering, K-Means, Ward Linkage*