

**PEMETAAN INDEKS PEMBANGUNAN MANUSIA
MENGUNAKAN ALGORITMA *K-MEANS***

THESIS

**Oleh :
UFIL HIDAYATUL LAILI
NIM. 200605220009**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PEMETAAN INDEKS PEMBANGUNAN MANUSIA
MENGUNAKAN ALGORITMA *K-MEANS***

THESIS

**Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
UFIL HIDAYATUL LAILI
NIM. 200605220009**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PEMETAAN INDEKS PEMBANGUNAN MANUSIA
MENGUNAKAN ALGORITMA K - MEANS**

THESIS

Oleh :
UFIL HIDAYATUL LAILI
NIM. 200605220009

Telah diperiksa dan disetujui untuk diuji:
Tanggal:

Pembimbing I,

Pembimbing II,

Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Dr. Fachrul Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Calyo Crysdiyan
NIP. 19740424 200901 1 008

**PEMETAAN INDEKS PEMBANGUNAN MANUSIA
MENGUNAKAN ALGORITMA *K - MEANS***

THESIS

Oleh :
UFIL HIDAYATUL LAILI
NIM. 200605220009

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 12 Desember 2022

Susunan Dewan Penguji

Tanda Tangan

Penguji Utama : Dr. Usman Pagalay, M.Si
NIP. 19650414 200312 1 001

()

Ketua Penguji : Dr. Totok Chamidy, M. Kom
NIP. 19691222 200604 1 001

()

Sekretaris Penguji : Dr. Fachrul Kurniawan ST., M.MT., IPM
NIP. 19771020 200912 1 001


()

Anggota Penguji : Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

()

Mengetahui dan Mengesahkan
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Calwo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Ufil Hidayatul Laili
NIM : 200605220009
Program Studi : Magister Informatika
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 08 November 2023

Yang membuat pernyataan,



Ufil Hidayatul Laili
NIM. 200605220009

MOTTO

وَإِلَىٰ رَبِّكَ فَأَرْجُ

Orang yang kuat tidak memamerkan harapannya. Orang yang kuat adalah mereka yang terus berusaha untuk mewujudkan harapannya. Dan hanya kepada Tuhanmulah hendaknya kamu berharap. (QS. Al Insyirah: 8)

HALAMAN PERSEMBAHAN

Thesis ini dipersembahkan untuk Bapak dan Ibu yang selalu mencurahkan segala daya dan upayanya selama menjalani pendidikan serta untuk putriku dengan doa dan supportnya.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Syukur alhamdulillah penulis haturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Muhammad Faisal, M.T dan Bapak Dr. Fachrul Kurniawan ST., M.MT ., IPM selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
2. Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
3. Ayahanda dan Ibunda tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
4. Anak penulis yang selalu memberikan semangat kepada penulis untuk menyelesaikan Thesis ini.
5. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materiil maupun moril.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

Wassalamu'alaikum Wr. Wb.

Malang, November 2023
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN.....	v
MOTTO.....	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR SIMBOL.....	xii
ABSTRAK.....	xiii
BAB I PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Perumusan Masalah	6
1.3. Tujuan Penelitian.....	6
1.4. Batasan Masalah.....	6
1.5. Manfaat Penelitian	6
BAB II STUDI LITERATURE	
2.1. Analisis Metode Clustering untuk Memetakan IPM	8
2.2. Landasan Teori.....	15
2.3. Kerangka Pemikiran	23
BAB III MODEL PENELITIAN	
3.1. Metode Pengumpulan Data	24
3.2. Pengolahan Data Awal.....	24
3.3. Metode yang diusulkan	29
BAB IV HASIL DAN PEMBAHASAN	
4.1. Implementasi <i>K-Means</i>	32
4.2. Tampilan Grafik <i>Clustering K - Means</i>	48
BAB V PENUTUP	
5.1. Simpulan	50
5.2. Penutup	51
DAFTAR PUSTAKA.....	52
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1 Kerangka Pemikiran Penelitian	23
Gambar 3.1 Metode yang diusulkan	29
Gambar 3.2 Algoritma <i>K-Means</i>	30
Gambar 4.1 Visualisasi <i>Clustering K-Means</i>	48

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terkait	12
Tabel 2.2 Nilai Maksimum dan Minimum Komponen IPM	19
Tabel 3.1 Data awal 2019.....	25
Tabel 3.2 Data Awal 2020.....	26
Tabel 3.3 Data awal 2021	27
Tabel 3.4 Data awal 2022	28
Tabel 4.1 Centroid awal	32
Tabel 4.2 Jarak setiap data ke centroid pada iterasi 1.....	34
Tabel 4.3 Anggota klaster 1 pada iterasi 1	35
Tabel 4.4 Anggota klaster 2 pada iterasi 1	36
Tabel 4.5 Anggota klaster 3 pada iterasi 1	36
Tabel 4.6 Anggota klaster 4 pada iterasi 1	37
Tabel 4.7 Anggota klaster 5 pada iterasi 1	37
Tabel 4.8 Centroid baru setelah iterasi 1	38
Tabel 4.9 Jarak setiap data ke centroid pada iterasi 2.....	39
Tabel 4.10 Anggota klaster 1 pada iterasi 2	40
Tabel 4.11 Anggota klaster 2 pada iterasi 2	41
Tabel 4.12 Anggota klaster 3 pada iterasi 3	41
Tabel 4.13 Anggota klaster 4 pada iterasi 4	42
Tabel 4.14 Anggota klaster 5 pada iterasi 5	42
Tabel 4.15 Centroid baru setelah iterasi 2.....	43
Tabel 4.16 Jarak setiap data ke centroid pada iterasi 4.....	43
Tabel 4.17 Daftar anggota Klaster tahun 2019.....	45
Tabel 4.18 Daftar anggota Klaster tahun 2020.....	46
Tabel 4.19 Daftar anggota Klaster tahun 2021	46
Tabel 4.20 Daftar anggota Klaster tahun 2022.....	47

ABSTRAK

Ufil Hidayatul Laili. 2023. **Pemetaan Indeks Pembangunan Manusia Menggunakan Algoritma K-Means**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Muhammad Faisal, M.T. (II) Dr. Fachrul Kurniawan ST., M.MT., IPM

Kata kunci : *Algoritma, Clustering, K - Means*

IPM diperkenalkan oleh United Nation Development Programe (UNDP) yang dikenal dengan istilah Human Development Index (HDI). IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. Pembangunan yang selama ini dilakukan oleh pemerintah merupakan upaya untuk mensejahterakan dan memakmurkan rakyat. Baik itu pembangunan dibidang infrastruktur, pendidikan, kesehatan, maupun ekonomi. Semua sasaran pembangunan tersebut adalah rakyat, harapannya dengan adanya pembangunan tersebut semua permasalahan yang ada pada rakyat diantaranya kemiskinan, pengangguran, putus sekolah, kesehatan dan kesenjangan sosial dapat teratasi. Program pembangunan yang dilakukan oleh Pemerintah Pusat maupun Pemerintah Daerah tidak dapat dilepaskan dari tinggi rendahnya nilai IPM. Salah satunya di Provinsi Jawa Timur, Khususnya Kabupaten Situbondo. IPM di Kabupaten Situbondo telah mengalami peningkatan dari tahun ke tahun. Pada tahun 2019 IPM Kabupaten Situbondo adalah sebesar 67,09 meningkat menjadi 67,38 pada tahun 2020 dan meningkat lagi menjadi 67,78 pada tahun 2021 dan pada tahun 2022 meningkat menjadi 68,25. Tujuan dilakukan pengelompokan Kabupaten/Kota di Provinsi Jawa Timur, melalui indikator – indikator IPM dari tahun 2019 hingga 2022. Selain sebagai bahan perencanaan program untuk tahun selanjutnya, pengelompokan yang dilakukan selama 2019 hingga 2022 juga sebagai evaluasi sasaran program Pemerintah Provinsi Jawa Timur yang sudah dilakukan pada tahun tersebut terutama berkaitan dengan program pembangunan kualitas hidup manusia.

ABSTRACT

Ufil Hidayatul Laili. 2023. **Mapping the Human Development Index Using the K-Means Algorithm**. Graduate Thesis. Master of Informatics, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim, Malang. Thesis Guide: (I) Dr. Muhammad Faisal, M.T. (II) Dr. Fachrul Kurniawan ST., M.MT., IPM

keywords: *Algorithm, Clustering, K-Means*

HDI was introduced by the United Nation Development Program (UNDP) which is known as the Human Development Index (HDI). HDI is an important indicator for measuring success in efforts to build the quality of human life. The development that has been carried out by the government is an effort to improve the welfare and prosperity of the people. Be it development in the fields of infrastructure, education, health or the economy. All the targets of development are the people, the hope is that with this development all the problems that exist among the people, including poverty, unemployment, school dropouts, health and social inequality can be resolved. Development programs carried out by the Central Government and Regional Governments cannot be separated from the high and low HDI values. One of them is in East Java Province, especially Situbondo Regency. HDI in Situbondo Regency has increased from year to year. In 2019 the HDI for Situbondo Regency was 67.09, increasing to 67.38 in 2020 and increasing again to 67.78 in 2021 and in 2022 increasing to 68.25. The aim of grouping districts/cities in East Java Province is through HDI indicators from 2019 to 2022. Apart from being a program planning material for the following year, the grouping carried out from 2019 to 2022 is also an evaluation of the program targets of the East Java Provincial Government which have been carried out in that year, it was mainly related to programs to develop the quality of human life.

ABSTRACT

أَطْرُوقَة. K-Means رسم خريطة لمؤشر التنمية البشرية باستخدام خوارزمية. Ufil Hidayatul Laili. 2023. برنامج الماجستير في المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، (I) Dr. Muhammad Faisal, M.T. (II) Dr. Fachrul Kurniawan ST., M.MT., IPM

م
ر

الكلمات المفتاحية: الخوارزمية، التجميع، K-Means

والذي يعرف باسم (UNDP) تم تقديم مؤشر التنمية البشرية من قبل برنامج الأمم المتحدة الإنمائي يعد مؤشر التنمية البشرية مؤشرا هاما لقياس النجاح في الجهود المبذولة لبناء (HDI) مؤشر التنمية البشرية نوعية الحياة البشرية. إن التطوير الذي نفذته الحكومة هو جهد لتحسين رفاهية ورخاء الشعب. سواء كان ذلك في مجالات البنية التحتية أو التعليم أو الصحة أو الاقتصاد. جميع أهداف التنمية هي الشعب، والأمل هو أنه مع هذه التنمية يمكن حل جميع المشاكل الموجودة بين الناس، بما في ذلك الفقر والبطالة والتسرب من المدارس والصحة وعدم المساواة الاجتماعية. ال يمكن فصل برامج التنمية التي تنفذها الحكومة المركزية والحكومات الإقليمية عن قيم مؤشر التنمية البشرية المرتفعة والمنخفضة. إحداها في مقاطعة جاوة الشرقية، وخاصة منطقة سيتوبونو. ارتفع مؤشر التنمية البشرية في سيتوبونو ريجنسي من سنة إلى أخرى. في عام 2019، بلغ مؤشر التنمية البشرية لمنطقة سيتوبونو 67.09، وزاد إلى 67.38 في عام 2020 وزاد مرة أخرى إلى 67.78 في عام 2021 وفي عام 2022 ارتفع إلى 68.25. الهدف من تجميع المناطق/المدن في مقاطعة جاوة الشرقية هو من خلال مؤشرات التنمية البشرية من عام 2019 إلى عام 2022. وبصرف النظر عن كونه مادة لتخطيط البرامج للعام التالي، فإن التجميع الذي تم تنفيذه من عام 2019 إلى عام 2022 هو أي تقييم الأهداف برنامج البرنامج. حكومة مقاطعة جاوة الشرقية التي تم تنفيذها في ذلك العام، كانت تتعلق بشكل أساسي ببرامج تطوير نوعية الحياة البشرية

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada dasarnya pembangunan adalah sebuah proses multidimensi mencakup beberapa perubahan penting dalam fondasi sosial, sikap dalam masyarakat dan lembaga nasional selain terus mempercepat pertumbuhan ekonomi sambil menghilangkan ketidaksetaraan pendapatan dan mengurangi kemiskinan (P & C, 2011). Sebagaimana firman Allah SWT dalam surat Al-A'raf ayat 96 yang berbunyi :

وَلَوْ أَنَّ أَهْلَ الْقُرَىٰ آمَنُوا وَاتَّقَوْا لَفَتَحْنَا عَلَيْهِم بَرَكَاتٍ مِّنَ السَّمَاءِ وَالْأَرْضِ¹

Dalam ayat tersebut diatas Allah telah menyatakan bahwa Dia akan menjamin bagi hamba-Nya yang beriman dan bertaqwa akan keberkahan yang datangnya dari langit dan bumi. Keberkahan tersebut bisa saja berupa kecukupan dan kesejahteraan bagi umat muslim yang jika dikaitkan dengan istilah sekarang kita sebut dengan IPM atau Indeks Pembangunan Manusia.

Selanjutnya sesuai sabda dari Nabi Muhammad SAW yang berbunyi :

وَأَيُّمَا أَهْلٌ عَرَصَتْ فِيهِمْ أَمْرٌ جَائِعٌ فَقَدْ بَرَأَتْ مِنْهُمْ ذِمَّةُ اللَّهِ تَعَالَى²

Dalam hadits ini dijelaskan bahwa sudah seharusnya bagi setiap manusia untuk saling tolong menolong kepada sesamanya. Apalagi bagi umat

¹ Artinya : “ Jika penduduk negeri beriman dan bertaqwa, niscaya kami membuka untuk mereka pintu keberkahan dari langit dan bumi ” (QS. Al-A'raf :96)

² Artinya : “ penduduk negeri manapun yang berada di pagi hari, sementara di tengah-tengah mereka ada orang yang kelaparan maka jaminan Allah telah lepas dari diri mereka ” (HR Ahmad, al-Hakim dan Abu Ya'la)

muslimin diwajibkan untuk mengatasi kemiskinan dan kelaparan bagi warga masyarakat lain yang kekurangan. Akan menjadi suatu hal yang aneh jika disuatu negara yang mayoritas muslim tetapi masih banyak warganya yang belum mencukupi kesejahteraannya.

Pembangunan yang selama ini dilakukan oleh pemerintah merupakan upaya untuk mensejahterakan dan memakmurkan rakyat. Baik itu pembangunan dibidang infrastruktur, pendidikan, kesehatan, maupun ekonomi. Semua sasaran pembangunan tersebut adalah rakyat, harapannya dengan adanya pembangunan tersebut semua permasalahan yang ada pada rakyat diantaranya kemiskinan, pengangguran, putus sekolah, kesehatan dan kesenjangan sosial dapat teratasi. Dengan kata lain, masyarakat dilibatkan dalam seluruh proses pembangunan. Untuk mengetahui keberhasilan pembangunan digunakanlah indeks yang dapat mengukurnya yaitu Indeks Pembangunan Manusia (IPM). IPM diperkenalkan oleh United Nation Development Programe (UNDP) yang dikenal dengan istilah Human Development Index (HDI). IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (Statistik, 2023).

IPM dibangun melalui tiga dimensi dasar yaitu umur panjang dan hidup sehat, pengetahuan serta standar hidup layak. Pada tahun 2010 IPM menggunakan metode perhitungan baru, ada 2 (dua) indikator yaitu Angka Melek Huruf pada metode lama diganti dengan Angka Harapan Lama Sekolah dan Produk Domestik Bruto (PDB) per kapita diganti dengan Produk Nasional

Bruto (PNB) per kapita. Dengan menggunakan metode penghitungan agregasi diubah dari rata-rata aritmatik menjadi rata-rata geometrik (Statistik, 2023).

Alasan yang dijadikan dasar perubahan metodologi penghitungan IPM yaitu Beberapa indikator sudah tidak tepat untuk digunakan dalam penghitungan IPM. Angka melek huruf sudah tidak relevan dalam mengukur pendidikan secara utuh karena tidak dapat menggambarkan kualitas pendidikan. Selain itu, karena angka melek huruf di sebagian besar daerah sudah tinggi, sehingga tidak dapat membedakan tingkat pendidikan antar daerah dengan baik. PDB per kapita tidak dapat menggambarkan pendapatan masyarakat pada suatu wilayah, kemudian penggunaan rumus rata-rata aritmatik dalam penghitungan IPM menggambarkan bahwa capaian yang rendah di suatu dimensi dapat ditutupi oleh capaian tinggi dari dimensi lain. Dengan begitu menggunakan rata-rata geometrik dalam menyusun IPM dapat diartikan bahwa capaian satu dimensi tidak dapat ditutupi oleh capaian di dimensi lain. Artinya, untuk mewujudkan pembangunan manusia yang baik, ketiga dimensi harus memperoleh perhatian yang sama besar karena sama pentingnya (Statistik, 2023).

Program pembangunan yang dilakukan oleh Pemerintah Pusat maupun Pemerintah Daerah tidak dapat dilepaskan dari tinggi rendahnya nilai IPM. Salah satunya di Provinsi Jawa Timur, Khususnya Kabupaten Situbondo. IPM di Kabupaten Situbondo telah mengalami peningkatan dari tahun ke tahun. Pada tahun 2019 IPM Kabupaten Situbondo adalah sebesar 67,09 meningkat menjadi 67,38 pada tahun 2020 dan meningkat lagi menjadi 67,78 pada tahun

2021 dan pada tahun 2022 meningkat menjadi 68,25. Kabupaten Situbondo menduduki Ranking ke 12 dari 38 Kabupaten/Kota di Jawa Timur yang IPM tertinggi adalah Kota Pacitan yaitu 69,37. Akan tetapi capaian IPM dari berbagai Kabupaten/Kota di Jawa Timur tidak menutup kemungkinan mempunyai kemiripan karakteristik nilai capaian IPM yang terkumpul dalam klaster yang sama. Sedangkan nilai capaian IPM yang berbeda akan terpisah dalam klaster yang berbeda.

Perlu dilakukan pengelompokan Kabupaten/Kota di Provinsi Jawa Timur, melalui indikator – indikator IPM dari tahun 2019 hingga 2022. Selain sebagai bahan perencanaan program untuk tahun selanjutnya, pengelompokan yang dilakukan selama 2019 hingga 2022 juga sebagai evaluasi sasaran program Pemerintah Provinsi Jawa Timur yang sudah dilakukan pada tahun tersebut terutama berkaitan dengan program pembangunan kualitas hidup manusia. Pengelompokan juga bertujuan untuk pemerataan pembangunan di Provinsi Jawa Timur. Dengan Pemerataan pembangunan terdapat jaminan bahwa semua penduduk dapat menikmati hasil-hasil pembangunan (Faris, 2017).

Analisis kelompok (*cluster analysis*) adalah pekerjaan mengelompokkan data (objek) yang di dasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya (Talakua, Leleury, & Talluta, 2017). Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan)

dengan objek dalam kelompok lain. Lebih besar kemiripannya (homogenitas) dalam kelompok dan lebih besar perbedaannya di antara kelompok yang lain (E, 2012).

Ada banyak metode pengelompokan yang sudah dikembangkan para ahli. Masing-masing metode mempunyai karakter, kelebihan dan kekurangan. Menurut struktur, pengelompokan dibagi menjadi dua, yaitu pengelompokan hierarki dan sekatan (*partitioning*). Dalam pengelompokan hierarki, satu data tunggal bisa dianggap sebuah kelompok, dua atau lebih kelompok kecil dapat bergabung menjadi sebuah kelompok besar dan begitu seterusnya hingga semua data dapat bergabung menjadi sebuah kelompok. Di sisi lain metode pengelompokan sekatan membagi set data ke dalam sejumlah kelompok yang tidak tumpang-tindih (tidak *overlap*) antara satu kelompok dengan kelompok yang lain. Artinya setiap data hanya menjadi anggota satu kelompok. Metode seperti *K-Means* masuk dalam kategori ini.

Beberapa penelitian telah menerapkan algoritma *K-Means* dalam melakukan pengelompokan, misalnya penelitian yang dilakukan oleh M. W. Talakua dkk. pada tahun 2017. Paper ini meneliti tentang Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014 menggunakan algoritma *K-Means* untuk menghasilkan perencanaan keputusan yang efektif. Hasil pengujian pada penelitian ini menunjukkan bahwa *K-Means* akan menghasilkan keputusan yang baik.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dibahas sebelumnya, peneliti merumuskan masalah yaitu mengelompokkan Indeks Pembangunan Manusia yang berkarakteristik sama di Jawa Timur menggunakan *K-Means* sehingga terklaster dan bermanfaat dalam perencanaan pembangunan manusia.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas tujuan penelitian ini adalah untuk melakukan pengelompokan Kabupaten/Kota berdasarkan capain Indeks Pembangunan Manusia dengan menggunakan algoritma *K-Means*.

1.4 Batasan Masalah

1. Data diambil dari web site resmi Badan Pusat Statistik
2. Data yang diambil Indeks Pembangunan Manusia Pada tahun 2019 – 2022

1.5 Manfaat Penelitian

1. Bagi Peneliti

Untuk menambah pengetahuan tentang metode yang digunakan dalam penelitian.

2. Bagi Pembaca

Untuk menambah pengetahuan tentang metode yang digunakan dalam penelitian, sebagai informasi, bahan pertimbangan dan acuan dalam melakukan penelitian selanjutnya.

3. Bagi Pemerintah

Menjadi bahan acuan dan masukan untuk mengevaluasi program-program yang menyangkut indikator-indikator IPM yang telah dilakukan di periode sebelumnya guna perencanaan dan pengambilan kebijakan agar tepat sasaran.

BAB II

STUDI LITERATUR

2.1 Analisis Metode *Clustering* Untuk Memetakan Indeks Pembangunan Manusia Dengan Algoritma *K-Means*

Penelitian terkait mengenai pembahasan analisis Metode *Clustering* Untuk Memetakan Indeks Pembangunan Manusia Dengan Algoritma *K-Means* telah banyak dilakukan pada penelitian sebelumnya. Menurut (Faris, 2017) Tujuan pembangunan yang dilakukan oleh pemerintah adalah untuk mensejahterakan rakyatnya. Salah satu alat ukur untuk mengukur keberhasilan pembangunan adalah Indeks Pembangunan Manusia (IPM). Tinggi rendahnya IPM tidak lepas dari peran pemerintah, termasuk Provinsi Banten. IPM Provinsi Banten mengalami peningkatan setiap tahunnya mulai tahun 2012 – 2015. Namun peningkatan IPM Provinsi Banten tersebut tidak diikuti oleh pemerataan pembangunan di Kabupaten/Kota yang ada di Provinsi Banten. Oleh karena itu perlu dilakukan pengelompokan Kabupaten/Kota di Provinsi Banten menurut indikator-indikator IPM. penelitian ini menggunakan dua metode pengelompokan yakni Fuzzy C-Means (FCM) dan KMedoids. Pengelompokan dilakukan selama tahun 2012-2015. Tujuannya untuk melihat pergerakan anggota cluster yang terbentuk tiap tahunnya.

Penelitian yang dilakukan (Rizal & Hakim, 2015) bertujuan pengelompokan variabel pembentuk Indeks Pembangunan Manusia di Kawasan Indonesia Timur pada tahun 2012 dengan menggunakan analisis

clustering, yaitu metode *K-Means* dan fuzzy c-means cluster. Variabel - variabel pembentuk IPM, terdiri dari angka harapan hidup, angka melek huruf, rata-rata lamanya sekolah, dan Purchasing Power Parity (paritas daya beli). Dari hasil analisis didapatkan fuzzy c-means menghasilkan nilai Sw/Sb lebih kecil dibandingkan metode c-means. Pada FCM jumlah kelompok optimum sebanyak enam kelompok dengan nilai Indeks Xie dan Beni 0,0006589. Kelompok 1 terdiri dari 32 kabupaten/kota, kelompok 2 terdiri dari 18 kabupaten/kota, kelompok 3 terdiri dari 26 kabupaten/kota, kelompok 4 terdiri dari 50 kabupaten/kota. Kelompok 5 terdiri dari 49 kabupaten/kota, dan kelompok 6 terdiri dari 44 kabupaten/kota.

Metode yang digunakan (Luthfi & Wijayanto, 2021) yaitu metode Algomerative menggunakan kemiripan jarak dengan Ward Method. Dari hasil ketiga metode tersebut akan dibandingkan untuk memperoleh metode yang terbaik dengan melihat ukuran validitas dari nilai Dunn Index (DN), Davies Bouldin Index (DB), Calinski-Harabasz Index (CH) serta untuk menentukan jumlah klaster/kelompok yang optimum dan yang terpenting dalam membandingkan untuk mendapatkan metode algoritma yang terbaik yaitu dengan memperoleh nilai rasio simpangan baku yang bertujuan untuk memperoleh nilai simpangan baku dalam kelompok (SW) yang minimum dan nilai simpangan baku antar kelompok (SB) yang maksimum. Model terbaik yang diperoleh yaitu menggunakan K-Medoids lebih baik dilihat dari perbandingan rasio simpangan baku kemudian diaplikasikan dalam analisis sentiment wilayah kabupaten/kota di Indonesia berdasarkan angka IPM

masing-masing wilayahnya sehingga didapatkan wilayah dengan angka IPM tertinggi dan wilayah dengan IPM terendah pada tahun 2019.

Mengacu pada penelitian (Talakua, Leleury, & Talluta, 2017) dalam penelitiannya membahas tentang pemanfaatan algoritma *K-Means* untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari lima ukuran Indeks Pembangunan Manusia (IPM). Lima ukuran tersebut adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Terdapat tiga cluster berdasarkan IPM yaitu: cluster 1: Kota Ambon, yang mempunyai angka IPM sangat maksimal. Cluster 2: MTB, Kepulauan Aru, SBB, SBT, MBD, dan Bursel, yang mempunyai angka IPM, AHH, AMH, RLS, dan PPK. Cluster 3: Malra, Malteng, Buru, Tual mempunyai angka IPM, AHH, AMH, RLS dan angka PPK. Berdasarkan angka Indeks Pembangunan Manusia, Angka Harapan Hidup, Angka Melek Huruf, Angka Rata-rata Lama Sekolah dan angka Pengeluaran Per Kapita, disimpulkan bahwa terdapat perbedaan yang signifikan pada tahun 2014.

(Rahmati, 2021) Penentuan jumlah cluster yang optimal dan metode pengelompokan terbaik dengan membandingkan Indeks Silhouette, Davis Bouldin dan Calinski Harabasz dari ketiga metode pengelompokan. Metode yang memiliki indeks optimal akan dipilih sebagai metode terbaik. Hasil yang didapat untuk pengelompokan data IPM Kabupaten/Kota tahun 2019 adalah terdapat 2 jumlah cluster optimal untuk metode *K-Means* dan Hierarchical dan 3 jumlah cluster untuk metode Fuzzy C-Means. Dengan membandingkan nilai

validasi antar ketiga metode, didapat bahwa metode *K-Means* adalah metode terbaik untuk pengelompokan data IPM Kabupaten/Kota tahun 2019.

Dalam penelitian (Muttaqin, 2020) bahwa penelitian ini dibagi menjadi tiga kelompok; daerah tinggi, daerah sedang dan daerah rendah. Kelompok pertama atau daerah rendah berisi 19 kota. Kelompok kedua atau area tengah berisi 381 kabupaten/kota. Kelompok ketiga atau daerah tinggi berisi 114 kabupaten/kota. Adapun beberapa cara atau kebijakan yang bisa diterapkan oleh pemerintah untuk mempercepat pertumbuhan kemakmuran di Indonesia dengan terfokus pada langkah-langkah percepatan seperti menarik anak-anak yang putus sekolah untuk kembali duduk di bangku sekolah, memaksa anak-anak yang terpaksa bekerja karena alasan ekonomi keluarga, untuk berhenti bekerja dan kembali ke bangku sekolah, menjadikan pemberantasan buta huruf sebagai sebuah gerakan yang berbasis desa/kelurahan dengan model intervensi by name by address, dan lain-lain.

Metode yang digunakan dalam penelitian (Santoso, 2019) adalah Analisis *K-Means* Cluster dan Analisis Diskriminan. Analisis *K-Means* Cluster digunakan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan indikator penyusun IPM tahun 2017. Hasil pengelompokkan dari Analisis *K-Means* Cluster, kemudian dianalisis lebih lanjut menggunakan Analisis Diskriminan untuk menguji keakuratan hasil dari pengelompokkan menggunakan Analisis *K-Means* Cluster. Analisis Diskriminan juga digunakan untuk mengetahui faktor yang paling berpengaruh dalam penyusunan nilai Indeks Pembangunan Manusia. Hasil analisis dari penelitian ini adalah

terbentuknya 2 kelompok (cluster) hasil dari Analisis *K-Means* Cluster. Karakteristik Cluster 1 berisi provinsi dengan umur harapan hidup (UHH) rendah, harapan lama sekolah (HLS) rendah, rata-rata lama sekolah (RLS) rendah, dan pengeluaran disesuaikan yang rendah. Karakteristik Cluster 2 berisi provinsi dengan umur harapan hidup (UHH) tinggi, harapan lama sekolah (HLS) tinggi, rata-rata lama sekolah (RLS) tinggi, dan pengeluaran disesuaikan yang tinggi. Model diskriminan 2 faktor yang adalah $Y = -11,642 + 1,409 RLS$. Jadi, indikator yang paling berpengaruh dalam penyusunan nilai IPM adalah rata-rata lama sekolah. Ketepatan hasil dalam mengklasifikasi kasus dari model diskriminan yang terbentuk adalah sebesar 91.20%. Jadi, dapat disimpulkan bahwa hasil pengelompokkan menggunakan metode *K-Means* Cluster mempunyai tingkat keakuratan yang tinggi.

Tabel 2.1 Rangkuman Penelitian Terkait

No	Peneliti	Tahun	Judul	Tujuan	Metode & Hasil	Rencana Penelitian
1	Muhamad Mirza Faris	2017	<i>Pengelompokkan Kabupaten/Kota Di Provinsi Banten Menurut Indikator Indeks Pembangunan Manusia Dengan Fuzzy C-Means Dan K-Medoids</i>	Untuk mengetahui pergerakan cluster yang terbentuk dalam pengelompokkan Kabupaten/Kota di Provinsi Banten berdasarkan indikator IPM tahun 2012 – 2015	penelitian ini menggunakan dua metode pengelompokkan yakni Fuzzy C-Means (FCM) dan KMedoids. Pengelompokkan dilakukan selama tahun 2012-2015.	Pada penelitian yang akan dilakukan lebih terfokus pada algoritma Implementasi <i>K-Means</i> .
2.	Annas Syaiful Rizal dan R.B	2015	<i>Metode K-Means Cluster Dan Fuzzy C-</i>	Pengelompokkan variabel pembentuk Indeks	fuzzy c-means menghasilkan nilai Sw/Sb lebih kecil	Yang akan dilakukan oleh peneliti

	Fajriya Hakim		<i>Means Cluster</i>	Pembangunan Manusia di Kawasan Indonesia Timur pada tahun 2012 dengan menggunakan analisis <i>clustering</i>	dibandingkan metode c-means. Pada FCM jumlah kelompok optimum sebanyak enam kelompok dengan nilai Indeks Xie dan Beni 0,0006589	hanya berfokus kepada nilai keakurasian dari algoritma klustering <i>K-Means</i> .
3.	Emir Luthfi, dan Arie Wahyu Wijayanto	2021	<i>Analisis perbandingan metode hirearchical, K-Means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia</i>	untuk dapat mengelompokkan suatu data observasi tertentu	Metode Hirearchical yang digunakan yaitu dengan metode Algotmerative menggunakan kemiripan jarak dengan Ward Method. Dari hasil ketiga metode tersebut akan dibandingkan untuk memperoleh metode yang terbaik.	Yang akan dilakukan peneliti hanya berfokus pada implementasi dari algoritma <i>K-Means</i> .
4.	M. W. Talakua, Z. A. Leleury, dan A. W. Talluta	2017	<i>Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014</i>	untuk mengukur keberhasilan sebuah negara. Salah satu aspek penting untuk mengukur tingkat pembangunan manusia adalah masyarakat yang unggul dari segi kuantitas dan kualitas.	untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari lima ukuran Indeks Pembangunan Manusia (IPM). Lima ukuran tersebut adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama	Pada Penelitian sebelumnya memang memiliki persamaan dalam perhitungan algoritma dengan peneliti, namun jika sebelumnya khusus di wilayah maluku sedangkan pada penelitian ini untuk implementa

					Sekolah (RLS), dan Pengeluaran Per Kapita (PPK).	si algoritma IPM di wilayah Jawa Timur.
5.	Rizqina Rahmati	2021	<i>Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering (Studi Kasus: Indeks Pembangunan Manusia tahun 2019)</i>	untuk mengelompokkan data atau objek yang didasarkan pada informasi yang ada untuk menggambarkan relasi yang terdapat antara objek tersebut.	Metode yang memiliki indeks optimal akan dipilih sebagai metode terbaik. Hasil yang didapat untuk pengelompokan data IPM Kabupaten/Kota tahun 2019 adalah terdapat 2 jumlah cluster optimal untuk metode K-Means dan Hierarchical dan 3 jumlah cluster untuk metode Fuzzy C-Means.	Yang akan dilakukan peneliti hanya berfokus pada implementasi dari algoritma K-Means.
6.	Muhammad Faishal Jundana Muttaqin	2020	<i>Analisis Kluster pada Kabupaten/ Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia.</i>	Dalam penelitian ini peneliti mengklasifikasi kabupaten atau kota di Indonesia berdasarkan IPM menjadi tiga kategori; daerah tinggi, sedang, dan rendah. peneliti menggunakan analisis kluster untuk penelitian ini.	Hasil penelitian ini dibagi menjadi tiga kelompok; daerah tinggi, daerah sedang dan daerah rendah. Kelompok pertama atau daerah rendah berisi 19 kota. Kelompok kedua atau area tengah berisi 381 kabupaten/kota. Kelompok ketiga atau daerah tinggi berisi 114 kabupaten/kota.	Peneliti hanya berfokus pada pengelompokan Indeks Pembangunan Manusia yang berkarakteristik sama di Jawa Timur menggunakan algoritma K-Means
7.	Muchlas Santoso	2019	<i>Analisis K-Means Cluster Indeks</i>	untuk menjelaskan karakteristik masing-	Metode yang digunakan dalam penelitian ini adalah Analisis K-	Yang akan dilakukan peneliti hanya

			<p><i>Pembangunan Manusia Di Indonesia</i></p>	<p>masing cluster yang diperoleh dari hasil pengelompokan provinsi di Indonesia berdasarkan indikator penyusunan IPM tahun 2017 menggunakan metode <i>K-Means Cluster</i>, menjelaskan model diskriminan 2 faktor yang terbentuk, serta untuk mengetahui ketepatan hasil dalam mengklasifikasi kasus dari model diskriminan yang terbentuk.</p>	<p><i>Means Cluster dan Analisis Diskriminan. Analisis K-Means Cluster</i> digunakan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan indikator penyusunan IPM tahun 2017. Hasil pengelompokan dari Analisis <i>K-Means Cluster</i>, kemudian dianalisis lebih lanjut menggunakan Analisis Diskriminan untuk menguji keakuratan hasil dari pengelompokan menggunakan Analisis <i>K-Means Cluster</i>.</p>	<p>berfokus pada implementasi dari algoritma <i>K-Means</i>.</p>
--	--	--	--	---	--	--

2.2 Landasan Teori

2.2.1 Indeks Pembangunan Manusia

Pembangunan manusia adalah suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Diantara banyak pilihan tersebut, pilihan yang terpenting adalah berumur panjang dan sehat, berilmu pengetahuan dan mempunyai akses terhadap sumber daya yang dibutuhkan agar dapat hidup secara layak.

a. Pengeluaran Per Kapita

Pengertian Pengeluaran per Kapita, menurut Badan Pusat Statistik Indonesia pengeluaran per kapita adalah biaya yang dikeluarkan untuk konsumsi anggota rumah tangga pada periode tertentu. (Statistik, 2023)

Sedangkan menurut (Halim, 2012), mendefinisikan pengeluaran per kapita secara keseluruhan bagi anggota rumah tangga yang termasuk dalam satu rumah tangga yaitu dengan memakai pengertian pengeluaran konsumsi rumah tangga. Jadi, Pengeluaran konsumsi rumah tangga yaitu pengeluaran yang dilakukan oleh rumah tangga untuk membeli barang-barang dan jasa-jasa kebutuhan hidup sehari-hari bagi anggota rumah tangga dalam suatu periode tertentu. Penghasilan rumah tangga atau uang masuk itu sebagian besar dibelanjakan lagi, yaitu untuk membeli segala hal yang diperlukan untuk hidup.

b. Umur Harapan Hidup Saat Lahir (UHH)

Umur Harapan Hidup (UHH) adalah perkiraan rata-rata tambahan umur seseorang yang diharapkan dapat terus hidup. UHH juga dapat didefinisikan sebagai rata-rata jumlah tahun yang dijalani oleh seseorang setelah orang tersebut mencapai ulang tahun yang ke-x. Ukuran yang umum digunakan adalah Umur Harapan Hidup saat lahir yang mencerminkan kondisi kesehatan pada saat itu. Sehingga pada umumnya ketika membicarakan UHH, yang

dimaksud adalah rata-rata jumlah tahun yang akan dijalani oleh seseorang sejak orang tersebut lahir. Umur Harapan Hidup merupakan alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk pada umumnya, dan meningkatkan derajat kesehatan pada khususnya (Bandung, 2019).

c. Harapan Lama Sekolah

Harapan Lama Sekolah didefinisikan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Diasumsikan bahwa peluang anak tersebut akan tetap bersekolah pada umur-umur berikutnya sama dengan peluang penduduk yang bersekolah per jumlah penduduk untuk umur yang sama saat ini. Harapan Lama Sekolah digunakan untuk mengetahui kondisi pembangunan system pendidikan di berbagai jenjang yang ditunjukkan dalam bentuk lamanya pendidikan (dalam tahun) yang diharapkan dapat dicapai oleh setiap anak.

d. Rata-Rata Lama Sekolah

Rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 15 tahun keatas dalam menjalani pendidikan formal.

Indikator komposit pembangunan manusia adalah alat ukur yang dapat digunakan untuk melihat pencapaian pembangunan manusia antar wilayah dan antar waktu. Indeks Pembangunan

Manusia (IPM) merupakan alat ukur yang dapat menunjukkan presentase pencapaian dalam pembangunan manusia dengan memperhatikan tiga faktor yaitu: kelangsungan hidup, pengetahuan, dan daya beli. Beberapa tahapan dalam penghitungan IPM dapat dijelaskan sebagai berikut :

- Tahap pertama penghitungan IPM adalah menghitung indeks masing-masing komponen IPM (Indeks Harapan Hidup = X1, Pengetahuan = X2 dan Standar Hidup Layak = X3)

$$\text{Indeks (Xi)} = (Xi - X_{\min}) / (X_{\max} - X_{\min})$$

Dimana :

Xi : Indikator komponen pembangunan manusia ke-i, i= 1,2,3

Xmin : Nilai minimum Xi

Xmaks : Nilai Maksimum Xi

Tabel 2.2 Nilai Maksimum dan Minimum Komponen IPM

Indikator Komponen IPM (=X(I))	Nilai maksimum	Nilai Minimum	Catatan
Angka Harapan Hidup	85	25	Sesuai standar global (UNDP)
Angka Melek Huruf	100	0	Sesuai standar global (UNDP)
Rata-rata lama sekolah	15	0	Sesuai standar global (UNDP)
Konsumsi per kapita yang disesuaikan	26.572.352	1.007.436	UNDP menggunakan PDB per kapita riil yang disesuaikan

- Tahapan kedua perhitungan IPM adalah menghitung rata-rata

sederhana dari masing-masing indeks X_i dengan rumus:

$$IPM = \{X_1 + X_2 + X_3\} / 3$$

dimana :

X_1 = Indeks Angka Harapan Hidup

$X_2 = 2/3$ (Indeks Melek Huruf) + $1/3$ (Indeks Rata-rata

Lama Sekolah) X_3 = Indeks Konsumsi perkapita yang disesuaikan

- Tahap ketiga adalah menghitung Reduksi Shortfall, yang digunakan untuk mengukur kecepatan perkembangan nilai IPM dalam suatu kurun waktu tertentu.

$$r = \{ (IPM_{t+n} - IPM_t) / (IPM_{ideal} - IPM_t) \times 100 \}^{1/n}$$

Dimana:

IPM_t = IPM pada tahun t

IPM_{t+n} = IPM pada tahun $t+n$

$IPM_{ideal} = 100$

2.2.2 Klastering

Klastering merupakan pekerjaan yang memisahkan data/vector ke dalam sejumlah kelompok (klaster) menurut karakteristiknya masing-masing. Data-data yang mempunyai kemiripan karakteristik akan berkumpul dalam klaster yang sama dan data-data dengan karakteristik berbeda akan terpisah dalam klaster yang berbeda. Tidak diperlukan

label kelas untuk setiap data yang diproses dalam klastering karena nantinya label baru bisa diberikan ketika klaster sudah terbentuk. Karena tidak adanya target label kelas untuk setiap data, maka klastering sering disebut juga pembelajaran tidak terbimbing (unsupervised learning) (Prasetyo, 2014).

Karena tidak ada label kelas yang digunakan dalam prosesnya, klastering sangat cocok untuk melakukan klastering data yang label kelasnya memang sulit didapatkan pada saat pembangkitan fitur. Pada klastering, segera setelah klastering terbentuk, maka label kelas untuk setiap data dapat diberikan dengan mengamati hasil klaster. Pekerjaan tidak terbimbing (unsupervised) seperti klastering juga sering digunakan untuk mengeksplorasi dan mengkaraktistikkan set data sebelum menjalankan pekerjaan yang terbimbing (supervised). Karena klastering tidak membutuhkan label kelas, yang perlu dicatat adalah bahwa kemiripan (similarity) harus didefinisikan berdasarkan pada atribut objek. Definisi kemiripan dan metode dalam data yang dikelompokkan berbeda tergantung dari algoritma klastering yang diterapkan. Algoritma klastering yang berbeda juga cocoknya digunakan pada jenis set data yang berbeda. Algoritma klastering yang bagus digunakan tergantung pada penerapan set data yang diproses.

2.2.3 K – Means

K-Means merupakan salah satu metode data klastering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih

klaster/kelompok. Metode ini mempartisi data ke dalam klaster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu klaster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain (Prasetyo, 2014).

Algoritma ini dapat menerima data tanpa ada label kategori. *K Means Clustering* memiliki objective yaitu meminimalisasi object function yang telah di atur pada proses clasterisasi. Dengan cara minimalisasi variasi antar 1 cluster dengan maksimalisasi variasi dengan data di cluster lainnya. *Clustering Algoritma (K-Means)* memiliki tujuan untuk meminimalisasikan fungsi objective yang telah di set dalam proses *clustering*. Tujuan tersebut dilakukan dengan cara meminimalikan variasi data yang ada didalam cluster dan memaksimalkan variasi data yang ada di cluster lainnya. Beberapa Karakteristik dari Algoritma *K-Means Clustering* yaitu sebagai berikut : (University, 2023).

Karakteristik dari *K-Means Cluster*:

- Cepat dalam proses *clustering*
- Sensitif terhadap nilai *centroid*
- Hasil dari *K-Means* selalu berubah ubah (dikarenakan tidak unik)
- Sulit meraih global optimum

Metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Inisialisasi: tentukan nilai K sebagai jumlah kluster yang diinginkan dan matrik ketidakmiripan (jarak) yang diinginkan. Jika perlu, tetapkan ambang batas perubahan fungsi objektif dan ambang batas perubahan posisi *centroid*.
2. Pilih K data dari set data X sebagai *centroid*.
3. Alokasikan semua data ke *centroid* terdekat dengan matrik jarak yang sudah ditetapkan (memperbarui kluster ID setiap data). Pengukuran jarak menggunakan *euclidean* yang diformulasikan oleh Persamaan 2.1

$$d(x, c) = \|x - c\|_2 = \sqrt{\sum r(x_i - c_i)^2} \quad 2.1$$

$x, c \in X$ dan x_i, c_i adalah nilai fitur ke- i dari x dan c , sedangkan r adalah jumlah fitur dalam vektor. Perhitungan data ke tiap *centroid* menggunakan Persamaan 2.2, 2.3 dan 2.4.

$$d(x_1, c_1) = \sqrt{\sum r(x_{1i} - c_{1i})^2} \quad 2.2$$

$$d(x_1, c_2) = \sqrt{\sum r(x_{1i} - c_{2i})^2} \quad 2.3$$

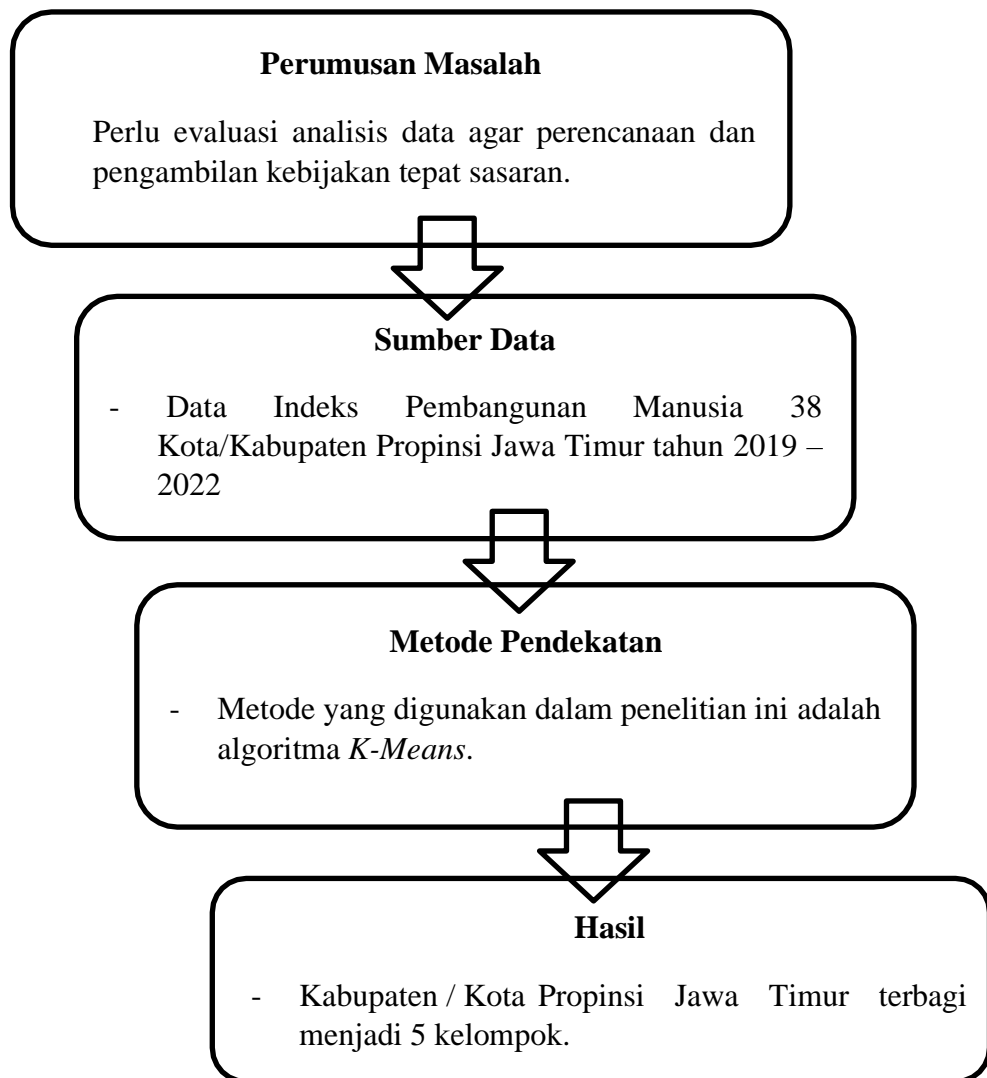
$$d(x_1, c_3) = \sqrt{\sum r(x_{1i} - c_{3i})^2} \quad 2.4$$

4. Hitung kembali *centroid* C berdasarkan data yang mengikuti kluster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu (a) perubahan fungsi objektif sudah di bawah ambang batas yang diinginkan; atau (b) tidak ada data yang berpindah kluster; atau (c)

perubahan posisi centroids sudah di bawah ambang batas yang diterapkan.

2.3 Kerangka Pemikiran

Kerangka pemikiran dalam penilitan ini ditunjukkan pada Gambar 2.1.



Gambar 2.1 Kerangka Pemikiran Penelitian

BAB III

METODE PENELITIAN

3.1. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari *website* resmi Badan Pusat Statistik Propinsi Jawa Timur (Statistik, 2023). Data Pengeluaran Per Kapita Jawa Timur dapat ditunjukkan pada Lampiran 1, data Umur Harapan Hidup saat Lahir Jawa Timur dapat ditunjukkan pada Lampiran 2, data Harapan Lama Sekolah Jawa Timur dapat ditunjukkan pada Lampiran 3 dan Rata – Rata Lama Sekolah Jawa Timur dapat ditunjukkan pada Lampiran 4.

3.2. Pengolahan Data Awal

Data yang didapatkan dari website resmi Badan Pusat Statistik Propinsi Jawa Timur yang terdiri dari berbagai parameter, data Indeks Pembangunan Manusia yang digunakan dalam penelitian ini merupakan gabungan dari data Pengeluaran per Kapita, Umur Harapan Hidup saat Lahir, Harapan lama Sekolah, dan Rata – rata lama Sekolah.

Data yang di ambil dari website resmi Badan Pusat Statistik Propinsi Jawa Timur tersebut mulai dari tahun 2019 - 2022, empat data digabung menjadi satu data. Adapun data awal dapat di tunjukkan pada Tabel 3.1 sampai dengan Tabel 3.4.

Tabel 3.1 Data awal 2019

Data ke i	W	X	Y	Z
1	9.033	71.77	12.62	7.28
2	9.883	72.65	13.72	7.21
3	9.865	73.59	12.25	7.28
4	10.891	73.95	13.15	8.07
5	10.861	73.39	12.45	7.29
6	11.146	72.54	12.88	8.01
7	10.270	72.45	13.17	7.27
8	9.274	69.94	11.80	6.22
9	9.525	68.99	13.22	6.18
10	12.264	70.54	12.78	7.13
11	10.665	66.55	13.27	5.71
12	10.097	68.97	13.14	6.12
13	10.972	67.00	12.34	5.77
14	10.381	70.17	12.31	7.11
15	14.609	73.98	14.91	10.25
16	12.860	72.43	12.61	8.49
17	11.533	72.27	13.00	8.53
18	12.200	71.44	12.85	7.63
19	11.650	71.22	13.14	7.80
20	11.779	72.49	14.00	7.96
21	11.468	72.16	12.69	6.98
22	10.265	71.36	12.36	7.09
23	10.499	71.26	12.20	6.81
24	11.572	72.27	13.47	7.89
25	13.295	72.61	13.72	9.29
26	8.718	70.11	11.59	5.66
27	8.760	67.96	12.08	4.55
28	8.834	67.45	13.63	6.40
29	9.082	71.22	13.19	5.46
30	12.440	73.96	14.97	9.92
31	13.851	73.60	14.31	10.10
32	16.666	73.15	15.41	10.17
33	12.280	70.19	13.57	8.69
34	13.393	71.40	13.60	9.11
35	13.710	73.21	13.83	10.24
36	16.040	72.75	14.39	11.13

37	17.854	74.13	14.79	10.47
38	12.870	72.54	14.12	9.06

Tabel 3.2 Data awal 2020

Data ke i	W	X	Y	Z
1	8.796	71.94	12.64	7.60
2	9.670	72.77	13.73	7.54
3	9.630	73.75	12.35	7.55
4	10.705	74.08	13.31	8.33
5	10.654	73.52	12.46	7.39
6	11.000	72.61	13.15	8.02
7	10.028	72.55	13.18	7.42
8	9.088	70.10	11.81	6.40
9	9.294	69.15	13.42	6.48
10	12.140	70.65	12.80	7.16
11	10.610	66.74	13.28	5.93
12	9.857	69.13	13.15	6.46
13	10.859	67.20	12.35	6.11
14	10.164	70.23	12.41	7.40
15	14.458	74.04	14.93	10.50
16	12.779	72.53	12.88	8.51
17	11.261	72.40	13.27	8.54
18	12.130	71.54	12.86	7.64
19	11.574	71.38	13.16	7.81
20	11.776	72.59	14.03	8.24
21	11.418	72.30	12.70	7.06
22	10.121	71.56	12.39	7.33
23	10.238	71.43	12.21	6.95
24	11.456	72.40	13.48	7.92
25	13.246	72.66	13.73	9.30
26	8.610	70.18	11.60	5.95
27	8.739	68.03	12.37	4.85
28	8.739	67.58	13.64	6.69
29	8.888	71.41	13.20	5.71
30	12.239	74.02	15.26	9.93
31	13.733	73.75	14.32	10.11
32	16.593	73.27	15.51	10.18

33	12.180	70.29	13.59	8.70
34	13.281	71.52	13.62	9.12
35	13.499	73.32	14.00	10.25
36	16.018	72.81	14.40	11.14
37	17.755	74.18	14.80	10.49
38	12.824	72.61	14.13	9.07

Tabel 3.3 Data awal 2021

Data ke i	W	X	Y	Z
1	8.887	72.07	12.65	7.61
2	9.851	72.85	13.74	7.55
3	9.743	73.86	12.47	7.56
4	10.807	74.16	13.32	8.34
5	10.757	73.61	12.63	7.50
6	11.127	72.65	13.44	8.08
7	10.163	72.61	13.24	7.43
8	9.203	70.21	11.88	6.67
9	9.410	69.28	13.43	6.49
10	12.217	70.72	13.10	7.42
11	10.690	66.89	13.29	5.94
12	9.996	69.24	13.16	6.62
13	10.969	67.36	12.36	6.12
14	10.297	70.25	12.58	7.41
15	14.578	74.06	14.94	10.72
16	12.844	72.59	12.95	8.64
17	11.394	72.49	13.57	8.55
18	12.172	71.60	12.87	7.78
19	11.658	71.50	13.17	7.82
20	11.833	72.65	14.04	8.36
21	11.459	72.41	12.83	7.26
22	10.221	71.72	12.68	7.38
23	10.380	71.56	12.22	7.18
24	11.510	72.49	13.77	8.04
25	13.280	72.67	13.77	9.56
26	8.673	70.22	11.73	5.96
27	8.790	68.07	12.38	4.86

28	8.804	67.67	13.65	6.70
29	9.000	71.56	13.33	5.92
30	12.359	74.04	15.27	10.15
31	13.816	73.86	14.33	10.35
32	16.663	73.36	15.75	10.41
33	12.245	70.35	13.60	8.95
34	13.354	71.60	13.63	9.33
35	13.610	73.39	14.01	10.47
36	16.095	72.83	14.41	11.37
37	17.862	74.18	14.81	10.50
38	12.887	72.65	14.16	9.31

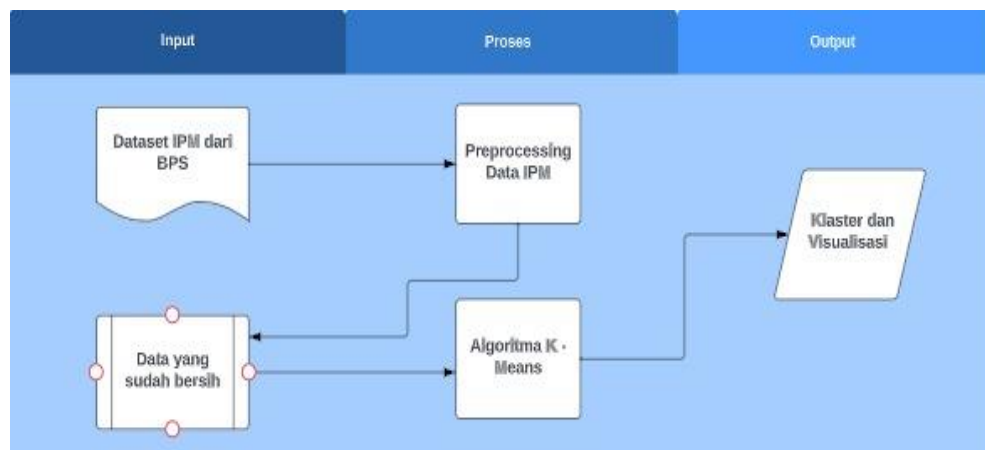
Tabel 3.4 Data awal 2022

Data ke i	W	X	Y	Z
1	9184	72.48	12.66	7.82
2	10199	73.20	13.76	7.77
3	10042	74.26	12.50	7.89
4	11162	74.54	13.33	8.65
5	11001	73.98	12.64	7.82
6	11565	72.97	13.61	8.23
7	10326	72.95	13.38	7.68
8	9466	70.61	12.02	6.87
9	9840	69.68	13.44	6.50
10	12320	71.06	13.11	7.66
11	10851	67.29	13.31	6.22
12	10263	69.62	13.18	6.63
13	11254	67.78	12.58	6.13
14	10726	70.55	12.76	7.42
15	14808	74.36	14.95	10.77
16	13051	72.93	12.96	8.97
17	11579	72.86	13.58	8.76
18	12349	71.95	13.07	8.12
19	11848	71.90	13.18	7.94
20	12031	72.97	14.05	8.66
21	11563	72.81	12.84	7.59
22	10323	72.16	12.84	7.43

23	10703	71.97	12.24	7.37
24	11648	72.86	14.01	8.33
25	13384	72.99	13.96	9.75
26	8971	70.54	11.91	5.97
27	8944	68.38	12.39	5.06
28	8967	68.03	13.67	6.88
29	9388	71.99	13.51	5.93
30	12762	74.34	15.44	10.45
31	14058	74.26	14.56	10.65
32	16897	73.75	15.76	10.69
33	12571	70.68	13.67	9.29
34	13803	71.96	13.64	9.67
35	14054	73.74	14.02	10.80
36	16503	73.13	14.43	11.67
37	18345	74.47	14.83	10.51
38	13094	72.97	14.40	9.63

3.3. Metode Yang Diusulkan

Metode yang diusulkan dapat ditunjukkan pada Gambar 3.1



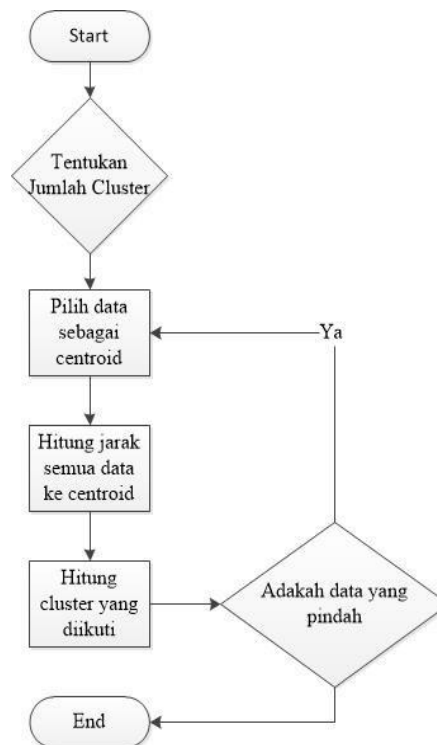
Gambar 3.1 Metode Yang Diusulkan

3.3.1 Selection

Dataset awal diambil dari data Indeks Pembangunan Manusia Jawa Timur mulai tahun 2019 sampai dengan 2022. Dari beberapa data diperbaiki dan dipilih (*selection*) dan digabung data-data yang diperlukan untuk penelitian ini (*selected data*).

3.3.2 Clustering.

Langkah-langkah pengklasteran dataset Indeks Pembangunan Manusia menggunakan algoritma *K-Means* dapat ditunjukkan pada gambar 3.2.



Gambar 3.2 Algoritma *K-Means*

Penentuan Klaster pada data Indeks Pembangunan Manusia Jawa Timur yang pertama dilakukan adalah sebagaimana tergambar dalam Algoritma *K-Means* diatas.

3.3.3 *Evaluasi*

Semakin kecil nilai yang di dapat maka hasil klastering yang didapat juga lebih baik. Secara esensial, DBI menginginkan nilai sekecil (non negatif ≥ 0) mungkin untuk menilai baiknya klaster yang didapat. Indeks tersebut didapat dari rata-rata semua indeks klaster.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Implementasi *K-Means*

Algoritma *K-Means* digunakan untuk mengelompokkan data yang ada. Data yang ada dikelompokkan menjadi 5 kelompok. Penentuan 5 kelompok ini didasarkan pada Data Indeks Pembangunan Manusia pada Badan Pusat Statistik yang dapat dikategorisasikan menjadi 5 klaster. Yakni klaster 1 menunjukkan sangat rendah, klaster 2 berarti rendah, klaster 3 adalah sedang, klaster 4 menunjukkan tinggi, serta klaster 5 yang menunjukkan sangat tinggi. Kategorisasi ini juga selaras dengan penelitian oleh Fauzi (Fauzi, 2017).

Fitur yang digunakan dalam klasterisasi *K-Means* adalah w , x , y dan z , nilai masing-masing dapat dilihat dalam Tabel mulai dari 3.1 sampai dengan 3.4 pada bab sebelumnya. Adapun langkah-langkah dari pengelompokan data adalah sebagai berikut:

1. Inisialisasi

Dilakukan pemilihan K data sebagai *centroid* awal sejumlah 5 *centroid*. Misalnya dipilih data ke 6, 12, 18, 24 dan 32. Data yang di pakai adalah data tahun 2019. Tabel 4.1 menunjukkan *centroid* awal yang dipilih.

Tabel 4.1 *Centroid* awal

<i>Centroid</i>	W	X	Y	Z
1	11,15	72,54	12,88	8,01
2	10,10	68,97	13,14	6,12
3	12,20	71,44	12,85	7,63
4	11,57	72,27	13,47	7,89
5	16,67	73,15	15,41	10,17

2. Iterasi 1

Menghitung jarak setiap data ke *centroid* terdekat. *Centroid* terdekat akan menjadi kluster yang diikuti oleh data tersebut. Berikut perhitungan jarak ke setiap *centroid* pada data ke-1:

$$d(x1, c1) = \sqrt{\sum r(x1i - c1i)^2}$$

$$= \sqrt{(9,033 - 11,15)^2 + (71,77 - 72,54)^2 + (12,62 - 12,88)^2 + (7,28 - 8,01)^2} = 2,38$$

$$d(x1, c2) = \sqrt{\sum r(x1i - c2i)^2}$$

$$= \sqrt{(9,033 - 10,10)^2 + (71,77 - 68,97)^2 + (12,62 - 13,14)^2 + (7,28 - 6,12)^2} = 3,25$$

$$d(x1, c3) = \sqrt{\sum r(x1i - c3i)^2}$$

$$= \sqrt{(9,033 - 12,20)^2 + (71,77 - 71,44)^2 + (12,62 - 12,85)^2 + (7,28 - 7,63)^2} = 3,21$$

$$d(x1, c4) = \sqrt{\sum r(x1i - c4i)^2}$$

$$= \sqrt{(9,033 - 11,57)^2 + (71,77 - 72,27)^2 + (12,62 - 13,47)^2 + (7,28 - 7,89)^2} = 2,79$$

$$d(x1, c5) = \sqrt{\sum r(x1i - c5i)^2}$$

$$= \sqrt{(9,033 - 16,67)^2 + (71,77 - 73,15)^2 + (12,62 - 15,41)^2 + (7,28 - 10,17)^2} = 8,74$$

Hasil perhitungan jarak setiap data ke *centroid* pada iterasi 1 seperti yang ditampilkan pada Tabel 4.2.

Tabel 4.2 Jarak setiap data ke *centroid* pada iterasi 1

Data ke-i	Jarak ke <i>centroid</i>					Klaster yang diikuti
	1	2	3	4	5	
1	2,38	3,25	3,21	2,79	8,74	C1
2	1,72	3,89	2,79	1,88	7,61	C1
3	1,92	4,85	3,25	2,55	8,05	C1
4	1,46	5,41	2,88	1,85	6,60	C1
5	1,23	4,69	2,42	1,78	7,13	C1
6	0,00	4,18	1,57	0,79	6,47	C1
7	1,19	3,67	2,23	1,48	7,40	C1
8	3,83	1,85	3,73	4,04	9,67	C2
9	4,32	0,58	3,92	4,23	9,43	C2
10	2,46	2,88	1,03	2,13	6,51	C3
11	6,45	2,52	5,49	6,19	10,20	C2
12	4,18	0,00	3,59	4,04	9,07	C2
13	6,00	2,33	4,99	5,82	9,95	C2
14	2,71	1,79	2,34	2,79	8,21	C2
15	4,82	8,10	4,83	4,45	2,28	C5
16	1,80	5,05	1,49	1,67	5,07	C3
17	0,71	4,33	1,40	0,79	5,97	C1
18	1,57	3,59	0,00	1,24	5,99	C3
19	1,45	3,21	0,68	1,11	6,30	C3
20	1,29	4,40	1,65	0,61	5,58	C4
21	1,16	3,61	1,23	1,21	6,75	C1
22	1,81	2,70	2,07	2,10	7,94	C1
23	1,99	2,60	2,01	2,23	7,95	C1
24	0,79	4,04	1,24	0,00	5,97	C4
25	2,64	5,82	2,47	2,26	3,91	C4

26	4,36	2,41	4,40	4,62	10,36	C2
27	6,27	2,53	5,83	6,29	11,49	C2
28	5,87	2,06	5,42	5,74	10,55	C2
29	3,55	2,56	3,82	3,64	9,40	C2
30	3,42	6,94	4,02	3,16	4,33	C4
31	3,85	7,26	3,95	3,54	3,06	C5
32	6,47	9,07	5,99	5,97	0,00	C5
33	2,78	3,61	1,79	2,34	5,79	C3
34	2,84	5,09	2,04	2,36	4,26	C3
35	3,59	6,96	3,63	3,33	3,35	C4
36	6,00	8,73	5,58	5,62	1,59	C5
37	7,56	10,41	7,14	7,16	1,69	C5
38	2,37	5,48	2,31	1,88	4,20	C4

Selanjutnya dihitung *centroid* yang baru untuk setiap klaster berdasarkan data yang bergabung pada setiap klasternya. Untuk klaster 1, ada 11 data yang bergabung ke dalamnya. Yaitu data ke 1,2,3,4,5,6,7,17,21,22,dan 23 sebagaimana pada Tabel 4.3.

Tabel 4.3 Anggota klaster 1 pada iterasi 1

Data Anggota	W	X	Y	Z
1	9,033	71,77	12,62	7,28
2	9,883	72,65	13,72	7,21
3	9,865	73,59	12,25	7,28
4	10,891	73,95	13,15	8,07
5	10,861	73,39	12,45	7,29
6	11,146	72,54	12,88	8,01
7	10,27	72,45	13,17	7,27
17	11,533	72,27	13,00	8,53

21	11,468	72,16	12,69	6,98
22	10,265	71,36	12,36	7,09
23	10,499	71,26	12,2	6,81
Jumlah	115,71	797,39	140,49	81,82
Rata-rata	10,52	72,49	12,77	7,44

Untuk kluster 2, ada 10 data yang bergabung ke dalamnya. Yaitu data ke 8, 9,11,12,13,14, 26, 27, 28 dan 29 sebagaimana pada Tabel 4.4

Tabel 4.4 Anggota kluster 2 pada iterasi 1

Data Anggota	W	X	Y	Z
8	9,274	69,94	11,8	6,22
9	9,525	68,99	13,22	6,18
11	10,665	66,55	13,27	5,71
12	10,097	68,97	13,14	6,12
13	10,972	67,00	12,34	5,77
14	10,381	70,17	12,31	7,11
26	8,718	70,11	11,59	5,66
27	8,76	67,96	12,08	4,55
28	8,834	67,45	13,63	6,4
29	9,082	71,22	13,19	5,46
Jumlah	96,31	688,36	126,57	59,18
Rata-rata	9,63	68,84	12,66	5,92

Untuk kluster 3, ada 6 data yang bergabung ke dalamnya. Yaitu data ke 10, 16, 18, 19, 33 dan 34 sebagaimana ditampilkan pada Tabel 4.5.

Tabel 4.5 Anggota klaster 3 pada iterasi 1

Data Anggota	W	X	Y	Z
10	12,264	70,54	12,78	7,13
16	12,86	72,43	12,61	8,49
18	12,2	71,44	12,85	7,63
19	11,65	71,22	13,14	7,8
33	12,28	70,19	13,57	8,69
34	13,393	71,4	13,6	9,11
Jumlah	74,65	427,22	78,55	48,85
Rata-rata	12,44	71,20	13,09	8,14

Untuk klaster 4, ada 6 data yang bergabung ke dalamnya. Yaitu data ke 20, 24, 25, 30, 35, dan 38 sebagaimana ditunjukkan pada Tabel 4.6.

Tabel 4.6 Anggota klaster 4 pada iterasi 1

Data Anggota	W	X	Y	Z
20	11,78	72,49	14,00	7,96
24	11,57	72,27	13,47	7,89
25	13,30	72,61	13,72	9,29
30	12,44	73,96	14,97	9,92
35	13,71	73,21	13,83	10,24
38	12,87	72,54	14,12	9,06
Jumlah	75,67	437,08	84,11	54,36
Rata-rata	12,61	72,85	14,02	9,06

Untuk klaster 5, ada 5 data yang bergabung ke dalamnya. Yaitu data ke 15, 31, 32, 36 dan 37 sebagaimana ditampilkan pada Tabel 4.7.

Tabel 4.7 Anggota klaster 5 pada iterasi 1

Data Anggota	W	X	Y	Z
15	14,609	73,98	14,91	10,25
31	13,851	73,6	14,31	10,1
32	16,666	73,15	15,41	10,17
36	16,04	72,75	14,39	11,13
37	17,854	74,13	14,79	10,47
Jumlah	79,02	367,61	73,81	52,12
Rata-rata	15,80	73,52	14,76	10,42

Rata-rata yang didapatkan dari 5 klaster tersebut adalah *centroid* baru yang didapatkan sebagaimana ditunjukkan pada Tabel 4.8.

Tabel 4.8 *Centroid* baru setelah iterasi 1

<i>Centroid</i>	W	X	Y	Z
1	10,52	72,49	12,77	7,44
2	9,63	68,84	12,66	5,92
3	12,44	71,20	13,09	8,14
4	12,61	72,85	14,02	9,06
5	15,80	73,52	14,76	10,42

3. Iterasi 2

Menghitung jarak setiap data ke *centroid* terdekat. *Centroid* terdekat akan menjadi klaster yang diikuti oleh data tersebut. Hasil perhitungan jarak setiap data ke *centroid* iterasi 2 seperti yang ditampilkan pada Tabel 4.9.

Tabel 4.9 Jarak setiap data ke *centroid* pada iterasi 2

Data ke-i	Jarak ke <i>centroid</i>					Klaster yang diikuti
	1	2	3	4	5	
1	1,67	3,29	3,59	4,37	7,96	C1
2	1,18	4,17	3,15	3,32	6,87	C1
3	1,39	4,97	3,71	3,79	7,17	C1
4	1,68	5,71	3,16	2,43	5,70	C1
5	1,03	4,92	2,90	2,99	6,29	C1
6	0,86	4,52	1,88	2,15	5,66	C1
7	0,50	3,94	2,65	3,09	6,65	C1
8	3,24	1,47	4,12	5,71	9,05	C2
9	3,88	0,65	4,16	5,77	8,96	C2
10	2,63	3,36	1,26	3,27	6,02	C3
11	6,21	2,59	5,55	7,43	9,97	C2
12	3,80	0,71	3,82	5,55	8,63	C2
13	5,77	2,30	5,10	7,11	9,66	C2
14	2,39	1,97	2,64	4,34	7,59	C2
15	5,61	8,67	4,49	2,74	1,30	C5
16	2,57	5,47	1,43	1,59	4,27	C3
17	1,52	4,73	1,46	1,68	5,15	C3
18	1,99	4,04	0,66	2,36	5,37	C3
19	1,78	3,68	0,86	2,43	5,66	C3
20	1,83	4,89	1,72	1,42	4,89	C4
21	1,11	3,94	1,84	2,81	6,07	C1
22	1,28	2,87	2,53	3,79	7,23	C1
23	1,49	2,76	2,52	3,92	7,27	C1
24	1,36	4,48	1,45	1,76	5,25	C1
25	3,47	6,34	2,10	0,82	3,08	C4
26	3,67	1,91	4,84	6,33	9,73	C2
27	5,70	1,93	6,16	7,93	11,06	C2

28	5,48	1,93	5,52	7,11	10,15	C2
29	2,79	2,55	4,30	5,36	8,81	C2
30	4,10	7,45	3,78	1,71	3,44	C4
31	4,67	7,79	3,61	1,81	2,03	C4
32	7,26	9,68	5,58	4,44	1,17	C5
33	3,25	4,17	1,26	2,74	5,29	C3
34	3,60	5,64	1,46	1,70	3,66	C3
35	4,43	7,47	3,25	1,66	2,32	C4
36	6,84	9,30	5,10	4,02	1,14	C5
37	8,35	11,00	6,79	5,63	2,14	C5
38	3,16	6,02	1,97	0,41	3,44	C4

Selanjutnya dihitung *centroid* yang baru untuk setiap klaster berdasarkan data yang bergabung pada setiap klasternya. Untuk klaster 1, ada 11 data yang bergabung ke dalamnya. Yaitu data ke 1, 2, 3, 4, 5, 6, 7, 21, 22, 23, dan 24 sebagaimana ditampilkan pada Tabel 4.10.

Tabel 4.10 Anggota klaster 1 pada iterasi 2

Data Anggota	W	X	Y	Z
1	9,033	71,77	12,62	7,28
2	9,883	72,65	13,72	7,21
3	9,865	73,59	12,25	7,28
4	10,891	73,95	13,15	8,07
5	10,861	73,39	12,45	7,29
6	11,146	72,54	12,88	8,01
7	10,27	72,45	13,17	7,27
21	11,468	72,16	12,69	6,98
22	10,265	71,36	12,36	7,09
23	10,499	71,26	12,2	6,81

24	11,572	72,27	13,47	7,89
Jumlah	115,75	797,39	140,96	81,18
Rata-rata	10,52	72,49	12,81	7,38

Untuk kluster 2, ada 10 data yang bergabung ke dalamnya. Yaitu data ke 8, 9, 11, 12, 13, 14, 26, 27, 28 dan 29 sebagaimana ditunjukkan pada Tabel 4.11.

Tabel 4.11 Anggota kluster 2 pada iterasi 2

Data Anggota	W	X	Y	Z
8	9,274	69,94	11,8	6,22
9	9,525	68,99	13,22	6,18
11	10,665	66,55	13,27	5,71
12	10,097	68,97	13,14	6,12
13	10,972	67	12,34	5,77
14	10,381	70,17	12,31	7,11
26	8,718	70,11	11,59	5,66
27	8,76	67,96	12,08	4,55
28	8,834	67,45	13,63	6,4
29	9,082	71,22	13,19	5,46
Jumlah	96,31	688,36	126,57	59,18
Rata-rata	9,63	68,84	12,66	5,92

Untuk kluster 3, ada 7 data yang bergabung ke dalamnya. Ditampilkan pada Tabel 4.12, yaitu data ke 10, 16, 17, 18, 19, 33 dan 34.

Tabel 4.12 Anggota kluster 3 pada iterasi 2

Data Anggota	W	X	Y	Z
10	12,264	70,54	12,78	7,13
16	12,86	72,43	12,61	8,49

17	11,533	72,27	13	8,53
18	12,2	71,44	12,85	7,63
19	11,65	71,22	13,14	7,8
33	12,28	70,19	13,57	8,69
34	13,393	71,4	13,6	9,11
Jumlah	86,18	499,49	91,55	57,38
Rata-rata	12,31	71,36	13,08	8,20

Untuk kluster 4, ada 6 data yang bergabung ke dalamnya. Yaitu data ke 20, 25, 30, 31, 35 dan 38 sebagaimana pada Tabel 4.13.

Tabel 4.13 Anggota kluster 4 pada iterasi 2

Data Anggota	W	X	Y	Z
20	11,78	72,49	14,00	7,96
25	13,30	72,61	13,72	9,29
30	12,44	73,96	14,97	9,92
31	13,85	73,60	14,31	10,10
35	13,71	73,21	13,83	10,24
38	12,87	72,54	14,12	9,06
Jumlah	77,95	438,41	84,95	56,57
Rata-rata	12,99	73,07	14,16	9,43

Untuk kluster 5, ada 4 data yang bergabung ke dalamnya. Yaitu data ke 15, 32, 36 dan 37 sebagaimana pada Tabel 4.14.

Tabel 4.14 Anggota kluster 5 pada iterasi 2

Data Anggota	W	X	Y	Z
15	14,609	73,98	14,91	10,25
32	16,666	73,15	15,41	10,17

36	16,04	72,75	14,39	11,13
37	17,854	74,13	14,79	10,47
Jumlah	65,17	294,01	59,50	42,02
Rata-rata	16,29	73,50	14,88	10,51

Rata-rata yang didapatkan dari 5 kluster tersebut adalah *centroid* baru yang didapat.

Tabel 4.15 *Centroid* baru setelah iterasi 2

<i>Centroid</i>	W	X	Y	Z
1	10,52	72,49	12,81	7,38
2	9,63	68,84	12,66	5,92
3	12,31	71,36	13,08	8,20
4	12,99	73,07	14,16	9,43
5	16,29	73,50	14,88	10,51

Hasil kluster yang diikuti pada iterasi 2 tidak sama dengan hasil kluster yang diikuti pada iterasi 1, maka proses kluster di lanjutkan lagi pada iterasi 3. Proses kluster berhenti pada iterasi 3 karena hasil yang diikuti pada iterasi 4 sama dengan hasil kluster yang diikuti pada iterasi 3. Adapun hasil perhitungan jarak setiap data ke *centroid* iterasi 4 seperti yang ditampilkan pada Tabel 4.16.

Tabel 4.16 Jarak setiap data ke *centroid* pada iterasi 4

Data ke-i	Jarak ke <i>centroid</i>					Kluster yang diikuti
	1	2	3	4	5	
1	1,58	3,29	3,31	5,30	8,44	C1
2	1,12	4,17	2,73	4,25	7,35	C1
3	1,31	4,97	3,32	4,61	7,66	C1

4	1,73	5,71	2,69	3,14	6,19	C1
5	1,03	4,92	2,51	3,82	6,76	C1
6	1,01	4,52	1,45	3,07	6,13	C1
7	0,45	3,94	2,26	4,05	7,12	C1
8	3,17	1,47	4,10	6,65	9,47	C2
9	3,84	0,65	4,19	6,70	9,36	C2
10	2,71	3,36	1,52	4,08	6,39	C3
11	6,20	2,59	5,79	8,22	10,27	C2
12	3,78	0,71	3,90	6,46	9,01	C2
13	5,77	2,30	5,37	7,90	9,97	C2
14	2,39	1,97	2,66	5,25	8,01	C2
15	5,74	8,67	4,36	1,82	1,77	C5
16	2,71	5,47	1,30	2,17	4,70	C3
17	1,68	4,73	1,04	2,56	5,62	C3
18	2,10	4,04	0,65	3,21	5,78	C3
19	1,89	3,68	0,72	3,34	6,09	C3
20	1,95	4,89	1,27	2,40	5,35	C3
21	1,16	3,94	1,56	3,73	6,50	C1
22	1,25	2,87	2,35	4,73	7,67	C1
23	1,46	2,76	2,39	4,86	7,71	C1
24	1,49	4,48	0,97	2,73	5,71	C3
25	3,62	6,34	1,97	0,86	3,55	C4
26	3,58	1,91	4,78	7,29	10,16	C2
27	5,62	1,93	6,24	8,86	11,43	C2
28	5,46	1,93	5,60	7,97	10,52	C2
29	2,67	2,55	4,10	6,34	9,25	C2
30	4,22	7,45	3,46	1,37	3,92	C4
31	4,80	7,79	3,45	0,84	2,54	C4
32	7,39	9,68	5,62	3,67	0,81	C5
33	3,37	4,17	1,54	3,36	5,66	C3

34	3,74	5,64	1,62	1,98	4,05	C3
35	4,58	7,47	3,13	0,79	2,81	C4
36	6,99	9,30	5,16	3,18	1,12	C5
37	8,48	11,00	6,83	4,81	1,69	C5
38	3,30	6,02	1,75	0,99	3,91	C4

Hasil klaster yang diikuti pada iterasi 4 sama dengan hasil klaster yang diikuti pada iterasi 3 berarti kondisi klaster tidak berubah maka proses iterasi berhenti. Adapun daftar keanggotaan klaster tahun 2019 dapat dilihat pada Tabel 4.17.

Tabel 4.17 Daftar Anggota Klaster Tahun 2019

No	Klaster	Anggota	Jumlah
1	Klaster 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Ngawi, Bojonegoro dan Tuban	10
2	Klaster 2	Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, Pamekasan dan Sumenep	10
3	Klaster 3	Banyuwangi, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Lamongan, Kota Probolinggo, dan Kota Pasuruan	9
4	Klaster 4	Gresik, Kota Kediri, Kota Blitar, Kota Mojokerto dan Kota Batu	5
5	Klaster 5	Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya	4

Adapun daftar keanggotaan klaster tahun 2020 dapat dilihat pada Tabel 4.18. Daftar keanggotaan klaster tahun 2021 dapat dilihat pada Tabel

4.19. dan Daftar keanggotaan klaster tahun 2022 dapat dilihat pada tabel 4.20.

Tabel 4.18 Daftar Anggota Klaster Tahun 2020

No	Klaster	Anggota	Jumlah
1	Klaster 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Jombang, Ngawi, Bojonegoro, Tuban dan Lamongan	12
2	Klaster 2	Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, Pamekasan dan Sumenep	10
3	Klaster 3	Banyuwangi, Nganjuk, Madiun, Kota Probolinggo, dan Kota Pasuruan	5
4	Klaster 4	Mojokerto, Magetan, Gresik, Kota Kediri, Kota Blitar, Kota Mojokerto dan Kota Batu	7
5	Klaster 5	Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya	4

Tabel 4.19 Daftar Anggota Klaster Tahun 2021

No	Klaster	Anggota	Jumlah
1	Klaster 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Ngawi, Bojonegoro, Tuban dan Sumenep	11
2	Klaster 2	Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, dan Pamekasan	9

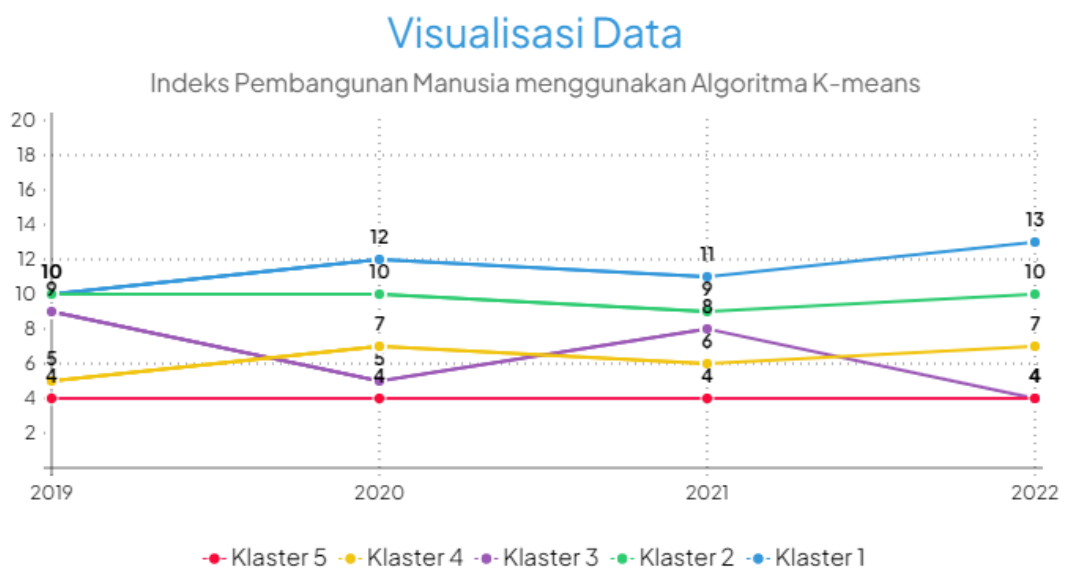
3	Klaster 3	Banyuwangi, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Lamongan, dan Kota Probolinggo	8
4	Klaster 4	Gresik, Kota Kediri, Kota Blitar, Kota Pasuruan, Kota Mojokerto dan Kota Batu	6
5	Klaster 5	Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya	4

Tabel 4.20 Daftar Anggota Klaster Tahun 2022

No	Klaster	Anggota	Jumlah
1	Klaster 1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Jombang, Magetan, Ngawi, Bojonegoro, Tuban, dan Lamongan	13
2	Klaster 2	Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, Pamekasan dan Sumenep	10
3	Klaster 3	Banyuwangi, Nganjuk, Madiun, dan Kota Probolinggo	4
4	Klaster 4	Mojokerto, Gresik, Kota Kediri, Kota Blitar, Kota Pasuruan, Kota Mojokerto dan Kota Batu	7
5	Klaster 5	Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya	4

4.2. Tampilan Grafik *Clustering K - Means*

Dari penjelasan dan tabel diatas dari tahun 2019 sampai dengan 2022 Indeks Pembangunan Manusia didapatkan bentuk Visualisasi Data pada Gambar 4.1 dibawah ini :



Gambar 4.1 Visualisasi *Clustering K-Means*

Pada Gambar 4.1, diperoleh bahwa warna Biru menunjukkan Klaster 1, kemudian warna Hijau menunjukkan Klaster 2, Klaster 3 yang berwarna Ungu, klaster 4 berwarna Kuning dan Klaster 5 berwarna Merah, klaster / kelompok tersebut dengan warna yang berbeda-beda agar lebih mudah dalam mengetahui kota mana saja dalam kelompok tersebut.

Diketahui pada tahun 2019 dalam klaster 1 terdapat 10 kota diantaranya Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Ngawi, Bojonegoro dan Tuban. Hasil penelitian tersebut diperoleh pengelompokan berdasarkan Indeks Pembangunan Manusia dengan menetapkan suatu obyek ke dalam cluster yang mempunyai *centroid* (mean) terdekat.

Begitupun juga pada klaster 2 pada tahun 2019 diperoleh 10 kota diantaranya Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, Pamekasan dan Sumenep. Kemudian pada klaster 3 diperoleh 9 kota yaitu diantaranya Banyuwangi, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Lamongan, Kota Probolinggo, dan Kota Pasuruan, selanjutnya pada klaster 4 yaitu diperoleh 5 kota yaitu diantaranya Gresik, Kota Kediri, Kota Blitar, Kota Mojokerto dan Kota Batu sedangkan yang terakhir pada klaster 5 yaitu diperoleh 4 kota yaitu diantaranya Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya. Mengacu pada Badan Pusat Statistik, klaster 1 dapat dikategorikan sebagai klasterisasi Indeks Pembangunan Manusia dengan kategori sangat rendah. Klaster 2 dapat dikategorikan rendah. Kategori sedang adalah kategori untuk klaster 3. Sementara klaster 4 dan 5 adalah kategori tinggi dan sangat tinggi.

BAB V

PENUTUP

5.1. Simpulan

Berdasarkan hasil penelitian dan pembahasan, diperoleh kesimpulan bahwa telah dilakukan pengelompokan Indeks Pembangunan Manusia dengan menggunakan algoritma *K-Means* sebanyak 5 *Centroid* (titik pusat). Pengelompokan menghasilkan pada tahun 2019 dalam Klaster 1 terdapat 10 kota diantaranya Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Ngawi, Bojonegoro dan Tuban. Kemudian pada klaster 2 pada tahun 2019 diperoleh 10 kota diantaranya Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sampang, Pamekasan dan Sumenep. Kemudian pada klaster 3 diperoleh 9 kota yaitu diantaranya Banyuwangi, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Lamongan, Kota Probolinggo, dan Kota Pasuruan, selanjutnya pada klaster 4 yaitu diperoleh 5 kota yaitu diantaranya Gresik, Kota Kediri, Kota Blitar, Kota Mojokerto dan Kota Batu sedangkan yang terakhir pada klaster 5 yaitu diperoleh 4 kota yaitu diantaranya Sidoarjo, Kota Malang, Kota Madiun dan Kota Surabaya. Mengacu pada Badan Pusat Statistik, klaster 1 dapat dikategorikan sebagai klasterisasi Indeks Pembangunan Manusia dengan kategori sangat rendah. Klaster 2 dapat dikategorikan rendah. Kategori sedang adalah kategori

untuk klaster 3. Sementara klaster 4 dan 5 adalah kategori tinggi dan sangat tinggi.

5.2. Saran

Penelitian ini dilakukan dengan menggunakan metode *K-Means*, dengan melakukannya terhadap data IPM Kabupaten/Kota Propinsi Jawa Timur, saran dalam pengembangan kedepannya diharapkan penelitian ini :

1. Sebaiknya variabel yang didapatkan dapat lebih banyak dan bervariasi agar hasil yang diperoleh lebih maksimal.
2. Sebaiknya dilakukan pengembangan dengan menggunakan optimasi untuk penentuan *centroid* agar hasil lebih maksimal.

DAFTAR PUSTAKA

- Bandung, S. k. (2019). *Analisis Pencapaian Sasaran : Kota Bandung*. Retrieved from Analisis Pencapaian Sasaran : Kota Bandung: <http://e-manekin.bandung.go.id/>
- E, P. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Faris, M. M. (2017). Pengelompokan Kabupaten/Kota Di Provinsi Banten Menurut Indikator Indeks Pembangunan Manusia Dengan Fuzzy C-Means Dan K-Medoids. *Universitas Islam Indonesia*, 3.
- Fauzi, F. (2017). K-Nearset Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah. *Indonesian Journal of Mathematics and Natural Sciences*, 40(2), 118-124.
- Halim, M. A. (2012). Teori Ekonomika Edisi 1. In M. A. Halim, *Teori Ekonomika Edisi 1* (p. 47). Tangerang: Jelajah Nusa.
- Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirearchical, *K-Means*, dan *k-medoids clustering* dalam pengelompokan indeks pembangunan manusia Indonesia. *Politeknik Statistika*, 3.
- Muttaqin, M. F. (2020). Analisis Klaster pada Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia = Cluster Analysis in Regencies/Cities of Indonesia Based on Human Development Index Indicators. *Universitas Indonesia Library*, 4.
- P, T. M., & C, S. S. (2011). *Pembangunan Ekonomi*. Jakarta: Erlangga.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- Rahmati, R. (2021). Analisis Cluster Dengan Algoritma *K-Means*, Fuzzy C-Means Dan Hierarchical *Clustering* (Studi Kasus: Indeks Pembangunan Manusia tahun 2019). *JIKO (Jurnal Informatika dan Komputer)*, 4.
- Rizal, A. S., & Hakim, R. F. (2015). METODE *K-MEANS CLUSTER* DAN *FUZZY C-MEANS CLUSTER* (Studi Kasus : Indeks Pembangunan Manusia di Kawasan Indonesia Timur tahun 2012). *Universitas Islam Indonesia Yogyakarta*, 3.
- Santoso, M. (2019). Analisis *K-Means* Cluster Indeks Pembangunan Manusia di Indonesia. *Unnes Library*, 5.
- Statistik, B. P. (2023, July 3). *Indeks Pembangunan Manusia*. Retrieved from Indeks Pembangunan Manusia: <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html>
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode *K-Means* Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Ilmu Matematika dan Terapan*, 121.

University, B. (2023). *School Of Information Systems*. Retrieved from School Of Information Systems: <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-K-Means/>