

**ANALISIS DATA WORLD DEVELOPMENT INDICATORS
MENGUNAKAN CLUSTER DATA MINING**

SKRIPSI



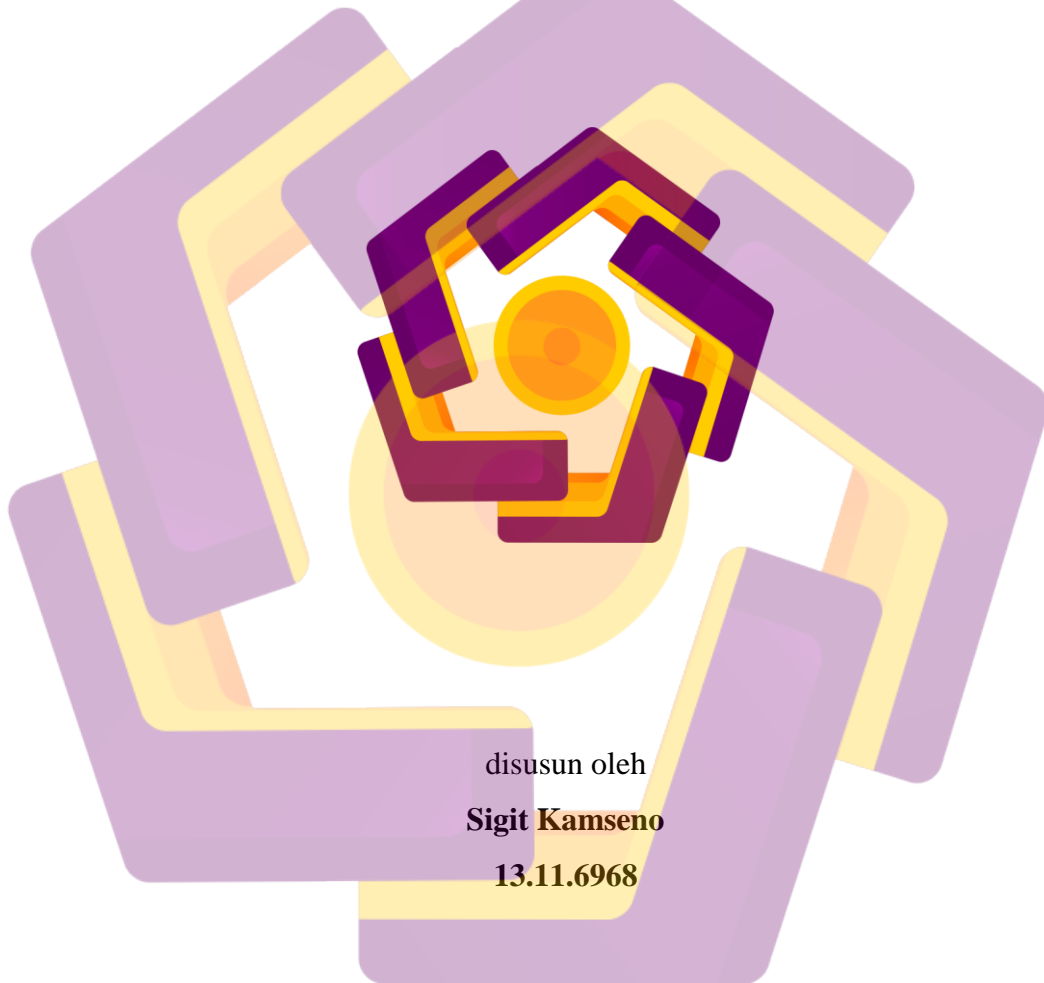
disusun oleh
Sigit Kamseno
13.11.6968

**PROGRAM SARJANA
PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2017**

**ANALISIS DATA WORLD DEVELOPMENT INDICATORS
MENGUNAKAN CLUSTER DATA MINING**

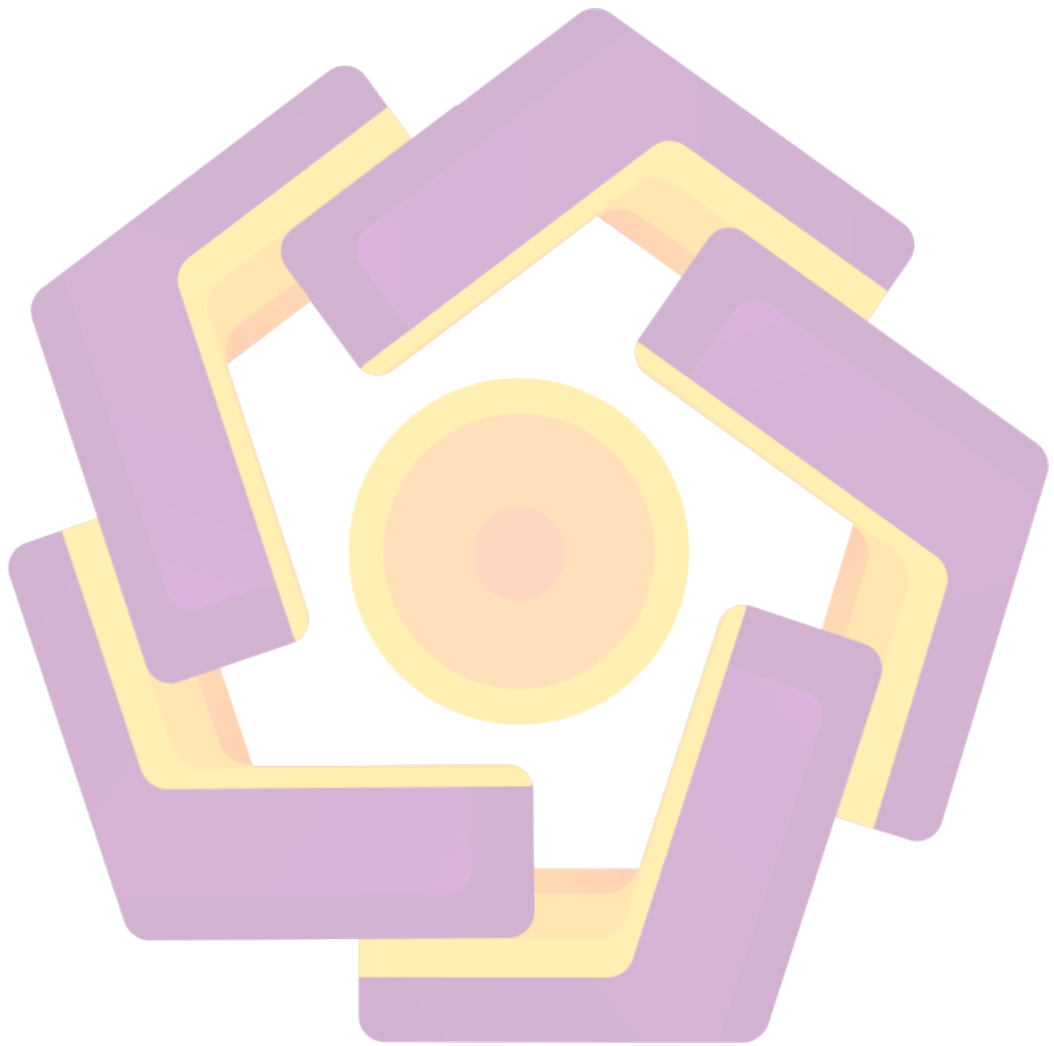
SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Teknik Informatika



disusun oleh
Sigit Kamseno
13.11.6968

**PROGRAM SARJANA
PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AMIKOM YOGYAKARTA
YOGYAKARTA
2017**



PERSETUJUAN

SKRIPSI

**ANALISIS DATA WORLD DEVELOPMENT INDICATORS
MENGUNAKAN CLUSTER DATA MINING**

yang dipersiapkan dan disusun oleh

Sigit Kamseno

13.12.6968

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Desember 2016

Dosen Pembimbing,



Barka Satya, M.Kom
NIK. 190302126

PENGESAHAN

SKRIPSI

**ANALISIS DATA WORLD DEVELOPMENT INDICATORS
MENGUNAKAN CLUSTER DATA MINING**

yang dipersiapkan dan disusun oleh

Sigit Kamseno

13.11.6968

telah dipertahankan di depan Dewan Penguji
pada tanggal 16 Januari 2017

Susunan Dewan Penguji

Nama Penguji

Sudarmawan, ST., M.T.
NIK. 190302035

Nila Feby Puspitasari, S.Kom., M.Cs
NIK. 190302161

Barka Satya, M.Kom
NIK. 190302126

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Januari 2017

KETUA STMIK AMIKOM YOGYAKARTA



Prof. Dr. M. Suvanto, M.M.
NIK. 190302001

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 26 Januari 2017



Sigit Kamseno

NIM. 13.11.6968

MOTTO

“ Ilmu tidak didapat dengan jasad yang santai. “



PERSEMBAHAN

Segala puja dan puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan keamanan, keselamatan, kelancaran, dan kebarokahan. Shalawat dan salam kepada nabi Muhammad SAW yang telah membawa dunia dari zaman biadab ke zaman yang lebih beradab. Dalam kesempatan kali ini, penulis juga tidak lupa mengucapkan rasa syukur dan terima kasih kepada :

1. Orang tua saya, Bapak Sarbini dan Ibu Darkini (Alm) yang selalu memberikan doa dan dukungannya sehingga saya dapat menyelesaikan skripsi ini tepat pada waktunya.
2. Kakak dan adik-adikku yang menyebarkan dan terkadang menyusahkan pula.
3. Pak Joko, Pak Hastari, Pak Ali, dan Pak Tony untuk segala bantuan, motivasi, kopi, rokok, dan sepiring nasi saat sedang bokek.
4. Bapak Barka Satya, M.Kom, atas bimbingannya selama dua semester.
5. Seluruh keluarga besar HMJTI yang memberi saya banyak pengalaman dan biji 1, biji 2, biji 3 dan biji - biji yang lain.
6. Seluruh keluarga besar FOSSIL yang selalu saya jadikan motivasi untuk berbuat lebih dari biasanya untuk lebih baik.
7. Adhatami atas bantuan dalam pemecahan rumus.

KATA PENGANTAR

Assalaamu'alaikum Warahmatullahi Wabarakaatu

Alhamdulillah, Puji Syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan skripsi ini yang diberi judul “ANALISIS DATA WORLD DEVELOPMENT INDICATORS MENGGUNAKAN CLUSTER DATA MINING”.

Laporan skripsi ini disusun sebagai syarat kelulusan di Sekolah Tinggi Manajemen Informatika dan Komputer “AMIKOM” Jurusan Teknik Informatika. Laporan ini dimaksudkan untuk memberikan kesempatan pada mahasiswa agar melihat, mengamati, membandingkan, menganalisis, serta menerapkan pengetahuan yang diperoleh diperguruan.

Dalam penyusunan skripsi ini, penulis tidak terlepas dari berbagai pihak yang telah rela membantu baik moril maupun materil yang dapat membuat penulis selalu optimis. Maka dari itu, sebagai rasa hormat penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua dan saudara - saudaraku yang telah memberikan dorongan moril maupun materil.
2. Bapak Prof. Dr. Mohammad Suyanto, MM selaku ketua Sekolah Tinggi Manajemen Informatika dan Komputer “AMIKOM” Yogyakarta.
3. Bapak Sudarmawan, S.T., M.T. selaku ketua jurusan S1 Teknik Informatika STMIK “AMIKOM” Yogyakarta.
4. Bapak Barka Satya, M.Kom selaku dosen pembimbing yang selalu sabar memberikan bimbingan, waktu dan arahan serta segala kemurahan hati kepada kami.

5. Seluruh Dosen dan karyawan STMIK AMIKOM Yogyakarta yang telah memberikan ilmunya kepada penulis.
6. All Crew HMJTI dan FOSSIL STMIK AMIKOM Yogyakarta.
7. Semua pihak yang telah membantu kelancaran penyusunan skripsi yang tidak dapat penulis sebutkan satu - persatu.

Penulis menyadari sepenuhnya bahwa laporan ini jauh dari sebuah kesempurnaan, itu semua karena keterbatasan penulis. Kritik dan saran yang bersifat membangun akan selalu penulis harapkan sehingga dapat lebih baik dan bermanfaat bagi penulis serta pihak - pihak yang membutuhkan.

Akhirnya dengan doa kepada Allah SWT, semoga laporan skripsi ini bermanfaat bagi semua pihak yang membutuhkan.

Wassalaamu'alaykum Warahmatullahi Wabarakatu

Yogyakarta, 26 Januari 2017

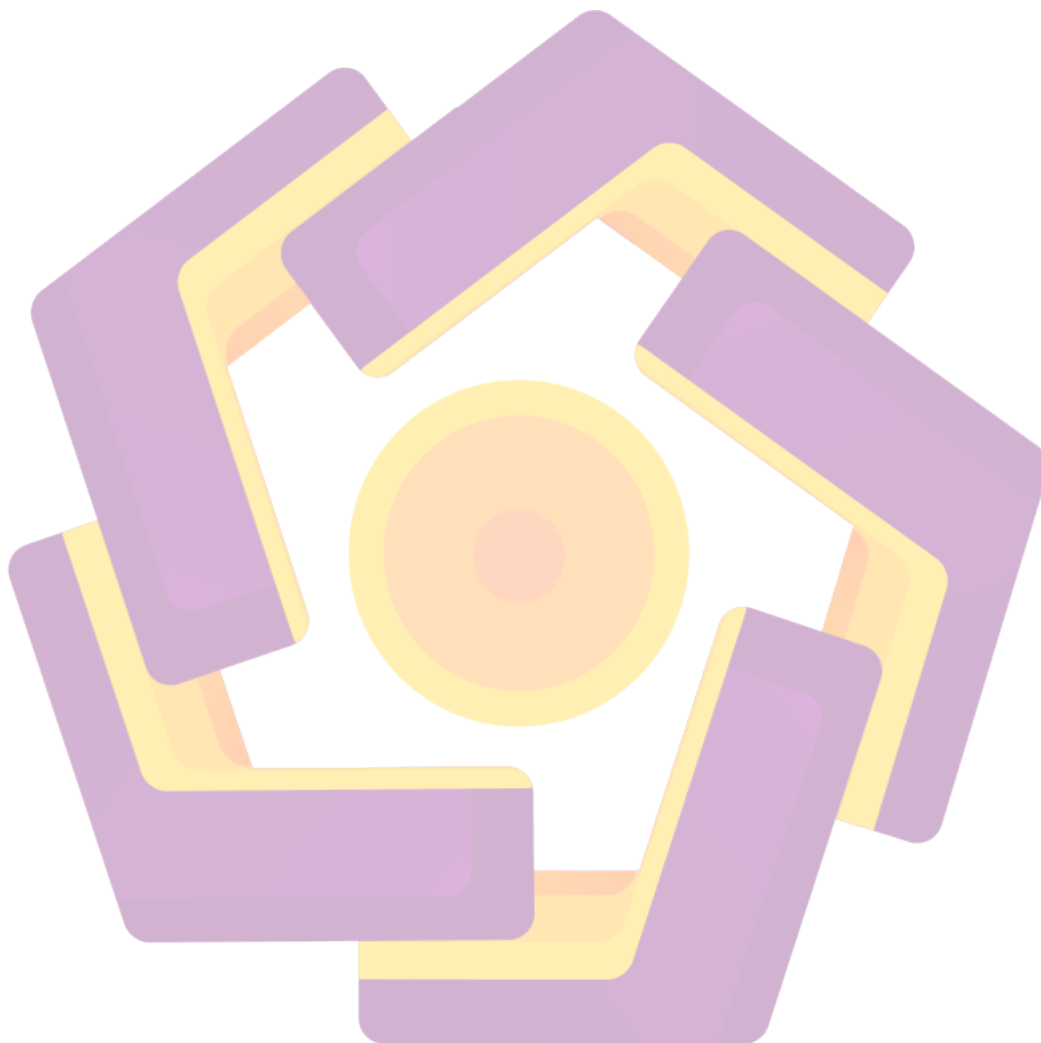
Penyusun

DAFTAR ISI

PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN.....	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Maksud dan Tujuan Penelitian.....	4
1.5. Manfaat Penelitian	4
1.6. Metodologi Penelitian	4
1.6.1. Pengumpulan Data	4
1.6.2. Metode Penelitian.....	5
1.6.3. Metode Pengembangan	5
1.6.4. Evaluasi <i>Cluster</i>	8
1.6.5. Tahapan Analisis	9
1.7. Sistematika Penulisan	9
BAB II LANDASAN TEORI	11
2.1. Tinjauan Pustaka	11
2.2. Dasar Teori.....	12
2.2.1. Pengertian Data Mining	12
2.2.2. Operasi Data Mining	12

2.2.3.	Analisis <i>Cluster</i>	13
2.2.4.	Dataset.....	14
2.2.5.	Algoritma DBSCAN.....	16
2.2.6.	Karakteristik DBSCAN.....	18
2.2.7.	<i>Sillhouette Coefficient</i>	19
2.3.	Konsep Pemodelan Sistem.....	21
2.3.1.	Teori <i>Flowchart</i>	21
2.3.2.	<i>Data Flow Diagram</i>	22
BAB III METODE PENELITIAN		25
3.1.	Tahap Pengembangan.....	25
3.1.1.	Planning.....	25
3.1.1.1.	Analisis Data.....	25
3.1.1.2.	Transformasi Data.....	26
3.1.1.3.	Analisis Model.....	27
3.1.2.	Designing.....	30
3.1.2.1.	Perancangan Alur Sistem.....	30
3.1.2.1.1.	Flowchart.....	30
3.1.2.1.2.	Diagram Konteks.....	31
3.1.2.1.3.	Data Flow Diagram.....	31
3.1.2.2.	Perancangan Antarmuka.....	32
3.1.2.2.1.	Halaman Utama.....	32
3.1.2.2.2.	Halaman Hasil.....	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1.	Coding.....	34
4.1.1.	Implementasi Program.....	34
4.1.2.	Pembuatan Tampilan.....	35
4.1.2.1.	Halaman Utama.....	35
4.1.2.2.	Halaman Hasil.....	36
4.2.	Testing.....	36
4.2.1.	Pengujian Hasil.....	36
4.2.1.1.	Nilai <i>Sillhouette Coefficient</i>	36
4.2.1.2.	Visualisasi Data.....	39

4.2.2. Perbandingan Hasil DBSCAN	40
4.2.3. Perhitungan Manual <i>Sillhouette Coefficient</i>	42
BAB V PENUTUP	48
5.1. Kesimpulan	48
5.2. Saran	49
DAFTAR PUSTAKA	50
LAMPIRAN	



DAFTAR TABEL

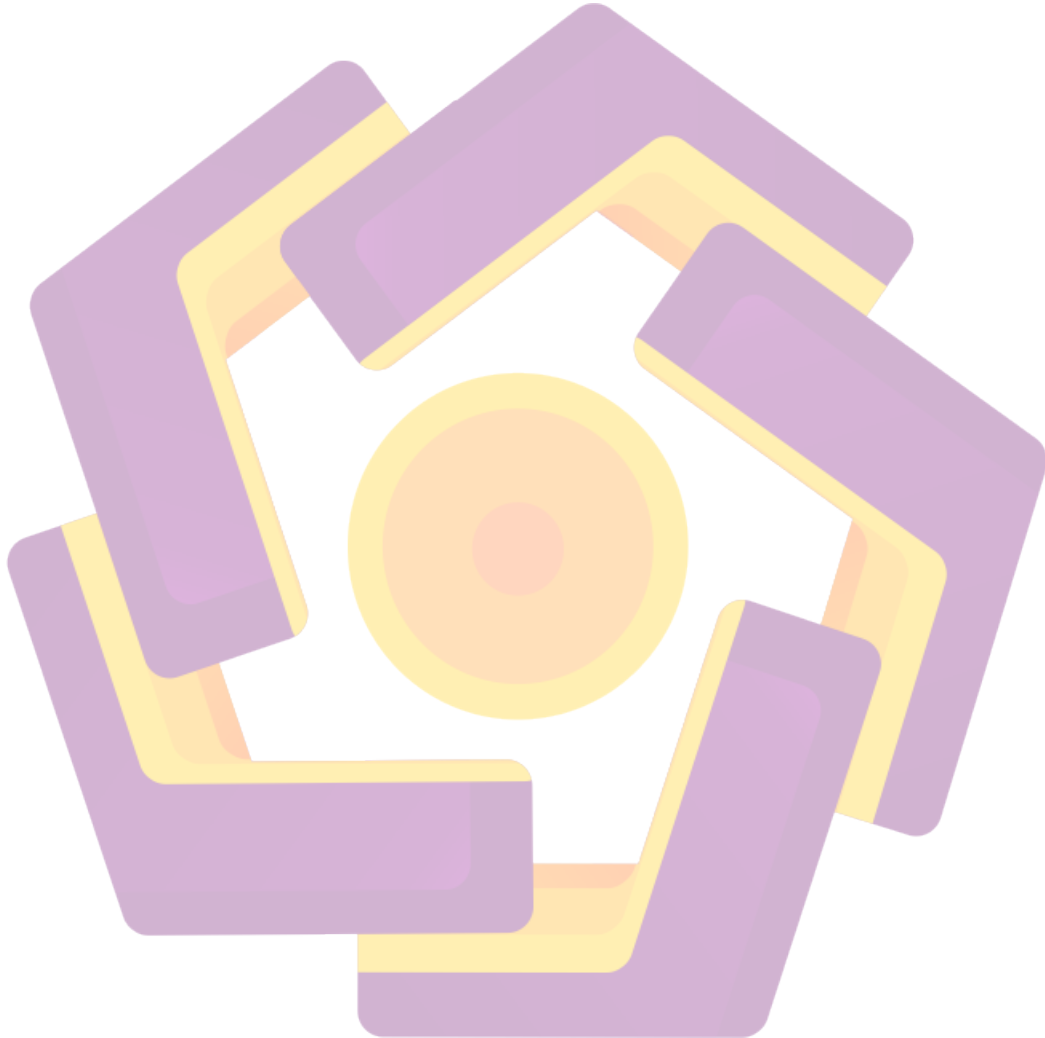
Tabel 2.1	Tipe Atribut.....	15
Tabel 2.2	Representasi Kauffman dan Rousseeuw (1990)	21
Tabel 2.3	Simbol - simbol <i>Flowchart</i>	21
Tabel 2.4	Simbol - simbol Data Flow Diagram	23
Tabel 3.1	Inisialisasi atribut <i>Income Group</i>	27
Tabel 3.2	inisialisasi atribut <i>Latest Trade Data</i>	27
Tabel 3.3	Contoh Data	28
Tabel 4.1	Hasil Klaster dan Nilai Silhouette.....	37
Tabel 4.2	Kemungkinan Klaster Terbaik.....	39
Tabel 4.3	Dataset.....	40
Tabel 4.4	Dataset Hasil Klaster Manual.....	41
Tabel 4.5	Hasil nilai SI untuk setiap data dalam klaster 1	44
Tabel 4.6	Hasil nilai SI untuk setiap data dalam klaster 2	46

DAFTAR GAMBAR

Gambar 1.1	Alur <i>Extreme Programming</i>	6
Gambar 2.1	Ilustrasi proses <i>Knowledge Discovery in database</i>	13
Gambar 2.2	Ilustrasi <i>Density-reachable</i> dan <i>Density-connected</i>	17
Gambar 3.1	Dataset.....	26
Gambar 3.2	Cluster dengan Epsilon 2	29
Gambar 3.3	Cluster dengan Epsilon $\sqrt{10}$	29
Gambar 3.4	<i>Flowchart</i>	30
Gambar 3.5	Diagram Konteks	31
Gambar 3.6	Data Flow Diagram	32
Gambar 3.7	Halaman Utama.....	33
Gambar 3.8	Halaman Hasil.....	33
Gambar 4.1	Halaman Utama.....	35
Gambar 4.2	Visualisasi Klaster.....	36
Gambar 4.3	Grafik Nilai <i>Silhouette Coefficient</i>	38
Gambar 4.4	Visualisasi 2 Klaster.....	40
Gambar 4.5	Visualisasi 3 Klaster.....	40
Gambar 4.6	Hasil Klaster Aplikasi	42

DAFTAR LAMPIRAN

Lampiran 1	Dataset.....	6
Lampiran 2	Main.py	10
Lampiran 3	Design.py	13



INTISARI

World Development Indicators merupakan sebuah database yang berisi tentang track records indikator - indikator yang mempengaruhi perkembangan suatu negara. Ada berbagai macam data yang tercatat dalam database tersebut, seperti nama negara, kode negara, sistem perdagangan yang digunakan, kategori pendapatan, survei - survei, dan masih banyak lagi. Data tersebut dihimpun oleh *World Bank* sebagai salah satu organisasi internasional yang berperan untuk membantu negara berkembang menjadi negara maju, khususnya dalam mengembangkan ekonomi.

Dalam penelitian ini akan dibuat kluster yang akan membagi negara - negara di dunia ke dalam sebuah kluster (kelompok) menggunakan algoritma DBSCAN. Kluster tersebut akan mewakili suatu negara termasuk dalam kluster negara maju, negara berkembang, atau negara yang tertinggal soal perkembangannya. Klusterisasi akan dilakukan dengan melakukan perbandingan indikator - indikator dimasing - masing negara yang kemudian indikator tersebut di transformasi menjadi nilai - nilai kuantitatif. Perbandingan nilai kuantitatif ditentukan dengan nilai eps (kedekatan), seberapa dekat nilai kuantitatif suatu negara dengan nilai kuantitatif negara lain. Setelah perbandingan nilai eps ditentukan pula minPts (minimum points) yang akan menentukan kedekatan - kedekatan points layak menjadi sebuah kluster. Kluster yang terbentuk akan mempengaruhi penamaan kluster, ditentukan berdasarkan rata - rata nilai kuantitatif negara di kluster tersebut lebih kecil dari kluster lain atau lebih besar dari kluster disekitarnya.

Evaluasi kluster dilakukan setelah kluster terbentuk dengan metode *Silhouette Index*. Metode ini dilakukan dengan mencari rata - rata jarak atau kemiripan kluster.

Kata Kunci: *Clustering, Dbscan, Silhouette Coefficient*

ABSTRACT

World Development Indicators is a database that contains records track indicators - indicators that influence the development of a country. There are various kinds of data recorded in the database, such as country name, country code, the trading system is used, the category of income, the survey - a survey, and many more. The data compiled by the World Bank as one of the international organizations whose role is to help developing countries into developed countries, particularly in developing economies.

In this study will be made cluster that will divide the country - a country in the world into a cluster (group) using algorithms DBSCAN. The cluster will represent a country included in the cluster of developed countries, developing countries or underdeveloped countries about development. Clustering will be done by doing a comparison of indicators - indicators dimasing - each country then these indicators in its transformation into a value - a quantitative value. Comparison of quantitative value determined by the value of eps (proximity), how close the quantitative value of a country with quantitative values of other countries. After Comparing the value eps is also determined minPts (minimum points) which will determine the proximity - proximity of points deserve to be a cluster. Clusters formed will influence the naming of clusters, determined based on the average - average value of quantitative countries in the cluster is smaller than the other clusters or larger than the surrounding cluster.

Cluster evaluation is done after cluster formed by the method Silhouette Index. This method is done by finding the average - average distance or similarity clusters.

Keyword: *Clustering, Dbscan, Sillhouette Coefficient*