ANALISA CLUSTERING SEKOLAH MENENGAH PERTAMA SEKOTA SEMARANG DENGAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING

Tugas Akhir disusun untuk mencapai gelar Kesarjanaan Komputer pada

Program Studi Teknik Informatika

Jenjang Program Strata-1

Oleh:

Harris Kristanto S.

15.01.53.0149

22205

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS STIKUBANK (UNISBANK)

SEMARANG

2021

PERNYATAAN KESIAPAN UJIAN TUGAS AKHIR

Saya, Harris Kristanto Santoso, dengan ini menyatakan bahwa Laporan Tugas Akhir yang berjudul:

ANALISA CLUSTERING SEKOLAH MENENGAH PERTAMA SEKOTA SEMARANG DENGAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING

adalah benar hasil karya saya dan belum pernah diajukan sebagai karya ilmiah, sebagian atau seluruhnya, atas nama saya atau pihak lain.

(Harris Kristanto Santoso)

NIM : 15.01.53.0149

Disetujui oleh Pembimbing

Kami setuju Laporan tersebut diajukan untuk Ujian Tugas Akhir

Semarang :

(Dra. Sulastri M. Kom.)

NIDN 0608126601

UNIVERSITAS STIKUBANK "UNISBANK" SEMARANG
FAKULTAS TEKNOLOGI INFORMASI
Rektorat Kampus Kendeng :
Jl.Kendeng V Bendan Ngisor Semarang Telp. (024) 8414970, Fax (024) 8441738
E-mail : fe@unisbank.ac.id
Kampus Mugas :
Jl.Tri Lomba Juang No. 1 Semarang 50241
Telp. (024) 8451976, 8311668, 8454746, Fax (024) 8443240 E-mail : info@unisbank.ac.id

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR/ SKRIPSI

Yang bertanda tangan di bawah ini, saya menyatakan bahwa TUGAS AKHIR / SKRIPSI dengan Judul :

PENENTUAN SEKOLAH MENENGAH PERTAMA TERBAIK SEKOTA SEMARANG DENGAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING

yang telah diuji di depan tim penguji pada tanggal 27 Januari 2021, adalah benar hasil karya saya dan dalam TUGAS AKHIR /SKRIPSI ini tidak terdapat keseluruhan atau sebagian tulisan orang lain yang saya ambil dengan cara menyalin, atau meniru dalam bentuk rangkaian kalimat atau simbol yang saya aku seolah-olah sebagai tulisan saya sendiri dan atau tidak terdapat bagian atau keseluruhan tulisan yang saya salin, tiru atau yang saya ambil dari tulisan orang lain tanpa memberikan pengakuan pada penulis aslinya.

Apabila saya melakukan hal tersebut diatas, baik sengaja maupun tidak, dengan ini saya menyatakan menarik TUGAS AKHIR / SKRIPSI yang saya ajukan sebagai hasil tulisan saya sendiri.

Bila kemudian terbukti bahwa saya ternyata melakukan tindakan menyalin atau meniru tulisan orang lain seolah-olah hasil pemikiran saya sendiri, berarti gelar dan ijazah saya yang telah diberikan oleh Universitas Stikubank (UNISBANK) Semarang batal saya terima.

Semarang , 3 Januari 2021

Yang Menyatakan

METERAI TEMPEL 62160ADF600780042 6000 ENAM RIBU RUPIAH

(HARRIS KRISTANTO SANTOSO)
NIM :15.01.53.0149

SAKSI 1
Tim Penguji

(SULASTRI, DRA, M.KOM)

SAKSI 2
Tim Penguji

(JEFFRI ALFA RAZAQ, M.Kom.)

SAKSI 3
Tim Penguji

(SETYAWAN WIBISONO, S.KOM, M.Cs)

HALAMAN PENGESAHAN

Tugas Akhir dengan judul:

Ditulis oleh :

ANALISA CLUSTERING SEKOLAH MENENGAH PERTAMA SEKOTA SEMARANG DENGAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING

NIM : 15.01.53.0149

Nama : Harris Kristanto Santoso

Telah dipertahankan di depan Tim Dosen Penguji Tugas Akhir dan diterima sebagai salah satu syarat guna menyelesaikan Jenjang Program Strata 1 Program Studi Teknik Informatika pada Fakultas Teknologi Informasi Universitas Stikubank (UNISBANK) Semarang.

Semarang, (27 Januari 2021)

Pembimbing / Penguji 1

(Dra. Sulastri, M.Kom)

NIDN : 0608126601

Penguji 2

(Jeffri Alfa Razaq, S.Kom., M.Kom.)

NIDN : 0611018401

Penguji 3

(Setyawan Wibisono, S.KOM, M.Cs)

NIDN : 0007067301

Mengetahui,

Universitas Stikubank (UNISBANK) Semarang

Fakultas Teknologi Informasi

Dekan

(Kristophorus Hadiono, S. Kom., Ph. D.)

NIDN : 0622027601

KATA PENGANTAR

Segala puji dan syukur bagi Tuhan Yang Maha Esa yang telah memberikan rahmat dan karuniaNya kepada penulis, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penyusunan skripsi ini dimaksudkan untuk memenuhi sebagian syarat-syarat guna mencapai gelar Sarjana Komputer di Universitas STIKUBANK (UNISBANK) Semarang.

Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun materil. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

1. Kedua orang tua, ayah tercinta Budi Santoso dan ibu tersayang Eny Kristianti yang telah memberikan dukungan baik moril maupun materil serta doa yang tiada henti-hentinya kepada penulis.
2. Bapak Dr. Safik Faozi, SH., M. Hum selaku Rektor Universitas STIKUBANK Semarang.
3. Bapak Kristophorus Hadiono, S. Kom., Ph. D., selaku Dekan Fakultas Teknologi Informasi Universitas STIKUBANK Semarang.
4. Bapak Dr. Drs. Eri Zuliarso, M. Kom. selaku Ketua Program Studi Fakultas Teknologi Informasi Universitas STIKUBANK Semarang.

5. Ibu Dra. Sulastri M. Kom, selaku dosen Pembimbing Skripsi yang telah berkenan memberikan tambahan ilmu dan solusi pada setiap permasalahan atas kesulitan dalam penulisan skripsi ini.
6. Seluruh Bapak/Ibu dosen Fakultas Teknologi Informasi yang telah memberikan pengetahuan yang sangat bermanfaat selama masa perkuliahan.
7. Seluruh staff dan karyawan Universitas STIKUBANK Semarang yang telah memberikan bantuan kepada penulis.
8. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak khususnya dalam bidang teknik informatika

Semarang, Januari 2021

Penulis,

(Harris Kristanto Santoso)

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG

Program Studi : Teknologi Informasi

Tugas Akhir Sarjana Komputer

Semester:10 tahun: 2021

ANALISA CLUSTERING SEKOLAH MENENGAH PERTAMA SEKOTA SEMARANG DENGAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING

Harris Kristanto Santoso

NIM : 15.01.53.0149

ABSTRAK

Pendidikan adalah hal yang penting dalam kehidupan kita, dengan adanya pendidikan, manusia dapat mempelajari ilmu pengetahuan yang selalu berkembang serta mencerdaskan generasi selanjutnya. Salah satu cara mendapatkan pendidikan adalah melalui sekolah. Dengan banyaknya pilihan sekolah, diperlukan informasi untuk menentukan predikat sebuah sekolah, salah satunya dengan nilai Ujian Negara (UN). Penelitian ini bertujuan melakukan clustering nilai UN tingkat SMP sekota Semarang dengan menggunakan algoritma *Hierarchical Agglomerative Complete* dan *Single Linkage Clustering,* kemudian membandingkan hasilnya. Data yang digunakan adalah nilai UN 217 SMP sekota Semarang dengan mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA.

Tools yang digunakan dalam penelitian ini adalah R Studio. Hasil clustering dibagi menjadi 3 cluster dengan tujuan menemukan SMP yang termasuk cluster unggul supaya para orang tua dapat memilih sekolah yang terbaik untuk anaknya. Setelah dilakukan clustering dengan metode *Complete Linkage*, didapatkan 10 SMP yang masuk golongan unggul yaitu SMP Negeri 2, SMP Negeri 21, SMP Karangturi, SMP PL Domenico Savio, SMP Maria Regina, SMP Mondial, SMP Negeri 30, SMP Negeri 1, SMP Negeri 5 dan SMP Negeri 9. Dengan metode *Average Linkage* didapatkan 10 SMP yang masuk ke cluster unggul yaitu SMP Daniel Creative, SMP Bina Bangsa School, SMP Singapore School, SMP Permata Bangsa, SMP Gandhi Memorial Intercontinental School, SMP Kristen Tri Tunggal, SMP Krista Mitra, SMP Maria Mediatrix, SMP Semesta Bilingual Boarding School dan SMP Islam Hidayatullah. Dari kedua metode ternyata ditemukan anggota clustering yang berbeda. Dalam penelitian kedepannya, diharapkan dapat melakukan clustering dengan metode lain serta melakukan validasi clustering.

ABSTRACT

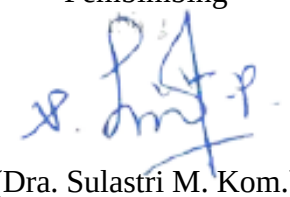
Education is a very important thing in our life. With education, human can learn knowledge that continuous to grow and enrich the life of a nation. Education can be achieved through school. In a city, there are many schools to choose to earn knowledge and how to determine predicate of a school, is by National Exam. The goal of this research is to clustering National Exam score of Junior High School level in Semarang with *Hierarchical Agglomerative Clustering Complete* and *Single Linkage Clustering*, afterward compare them. The dataset is National Exam score from 217 Junior High School in Semarang from 4 subjects specifically Indonesian, English, Matemathics and Science

R Studio is utilized in this research to cluster National Exam results. The results of the clustering are divided into 3 clusters in order to find which Junior High Schools are assessed as "Best". The result from using *Complete Linkage* method are SMP Negeri 2, SMP Negeri 21, SMP Karangturi, SMP PL Domenico Savio, SMP Maria Regina, SMP Mondial, SMP Negeri 30, SMP Negeri 1, SMP Negeri 5 dan and SMP Negeri 9. While using *Average Linkage*, the results are SMP Daniel Creative, SMP Bina Bangsa School, SMP Singapore School, SMP Permata Bangsa, SMP Gandhi Memorial Intercontinental School, SMP Kristen Tri Tunggal, SMP Krista Mitra, SMP Maria Mediatrix, SMP Semesta Bilingual Boarding School dan SMP Islam Hidayatullah. The results from both shows significant difference in each cluster members. In future, new researchs are expected to use different clustering methods and add clustering validation.

Kata Kunci

Hierarchical Agglomerative Clustering, Average Linkage, Complete Linkage, Analisa Clustering.

Pembimbing



(Dra. Sulastri M. Kom.)

DAFTAR ISI

HALAMAN JUDUL i
PERNYATAAN KESIAPAN UJIAN TUGAS AKHIR ii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR iii
HALAMAN PENGESAHAN iv
KATA PENGANTAR v
ABSTRAK vii
DAFTAR ISI ix
DAFTAR TABEL xi
DAFTAR GAMBAR xiii
DAFTAR RUMUS xvi
DAFTAR LAMPIRAN xvii
BAB I 1
PENDAHULUAN 1
1.1. LATAR BELAKANG PENELITIAN 1
1.2. PERUMUSAN MASALAH 4
1.3. TUJUAN DAN MANFAAT PENELITIAN 4
1.3.1. TUJUAN 4
1.3.2. MANFAAT 5
1.4. METODE PENELITIAN 6
1.4.1. PERUMUSAN OBYEK PENELITIAN 6
1.4.2. METODE PENGUMPULAN DATA 7
1.4.3. METODE ANALISIS 7
1.5. SISTEMATIKA PENULISAN 8
BAB II 10
TINJAUAN PUSTAKA 10
2.1. DATA MINING 10
2.2. CLUSTERING 11
2.3. PENELITIAN YANG TERKAIT 15

2.4. PERBEDAAN DENGAN PENELITIAN TERDAHULU 17
BAB III 18
METODE PENELITIAN..... 18
3.1. TOOLS YANG DIGUNAKAN 18
3.2. PENENTUAN VARIABEL..... 18
3.3. TEKNIK PENGUMPULAN DAN ANALISA DATA..... 19
3.4. PENGUKURAN JARAK..... 19
3.5. DATA CLEANSING 20
3.6. METODE CLUSTERING 23
a. PERHITUNGAN EUCLIDEAN DISTANCE 24
b. PERHITUNGAN AVERAGE LINKAGE..... 26
c. PERHITUNGAN COMPLETE LINKAGE..... 27
d. SOFTWARE R-STUDIO..... 28
BAB IV 29
HASIL DAN PEMBAHASAN..... 29
4.1. PROSES AWAL 29
4.2 CLUSTERING UNTUK SMP SE-KOTA SEMARANG..... 35
4.2. Clustering untuk SMP Swasta 38
4.3. Clustering untuk SMP Negeri 42
4.4. Clustering untuk 4 nilai mata pelajaran..... 45
BAB V..... 58
KESIMPULAN DAN SARAN..... 58
5.1. Kesimpulan..... 58
5.2. Saran 60
DAFTAR PUSTAKA 61

DAFTAR TABEL

TABEL 3.1 DATA AWAL……21

TABEL 3.2 DATA CLEANSING……22

TABEL 3.3 DATA SAMPEL NAMA SEKOLAH……...23

TABEL 3.4 DATA SAMPEL NILAI UN……23

TABEL 3.5 DATA MENTAH……...25

TABEL 3.6 MATRIKS DIAGONAL NILAI BAHASA INDONESIA……27

TABEL 4.1 HASIL CLUSTERING COMPLETE LINKAGE UN SEKOTA SEMARANG……37

TABEL 4.2 HASIL CLUSTERING AVERAGE LINKAGE UN SEKOTA SEMARANG……39

TABEL 4.3 HASIL CLUSTERING COMPLETE LINKAGE UN SMP SWASTA……...41

TABEL 4.4 HASIL CLUSTERING AVERAGE LINKAGE UN SMP SWASTA……...42

TABEL 4.5 HASIL CLUSTERING COMPLETE LINKAGE UN SMP NEGERI……44

TABEL 4.6 HASIL CLUSTERING AVERAGE LINKAGE UN SMP NEGERI……46

TABEL 4.7 HASIL CLUSTERING COMPLETE LINKAGE UN BAHASA INDONESIA……..47

TABEL 4.8 HASIL CLUSTERING AVERAGE LINKAGE UN BAHASA INDONESIA……..49

TABEL 4.9 HASIL CLUSTERING COMPLETE LINKAGE UN BAHASA INGGRIS……50

TABEL 4.10 HASIL CLUSTERING AVERAGE LINKAGE UN BAHASA INGGRIS……52

TABEL 4.11 HASIL CLUSTERING COMPLETE LINKAGE DATA UN MATEMATIKA……53

TABEL 4.12 HASIL CLUSTERING AVERAGE LINKAGE UN MATEMATIKA……55

TABEL 4.13 HASIL CLUSTERING COMPLETE LINKAGE UN IPA…….56

TABEL 4.14 HASIL CLUSTERING AVERAGE LINKAGE UN IPA……...58

DAFTAR GAMBAR

GAMBAR 4.1 HALAMAN UTAMA R STUDO CLOUD……..29

GAMBAR 4.2 HALAM LOGIN R STUDIO……30

GAMBAR 4.3 PROSES MEMULAI PROJECT R STUDIO……...30

GAMBAR 4.4 PROSES UPLOAD DATA……...31

GAMBAR 4.5 PROSES PEMILIHAN DATA……31

GAMBAR 4.6 DATA SUDAH TERDAFTAR DI REPOSITORY R STUDIO.32

GAMBAR 4.7 PROSES INSTALL PACKAGE……..32

GAMBAR 4.8 PEMILIHAN PACKAGE……33

GAMBAR 4.9 BARIS PROGRAM CLUSTERING……33

GAMBAR 4.10 RINGKASAN NILAI UN SMP SEKOTA SEMARANG…….35

GAMBAR 4.11 CLUSTERING COMPLETE LINKAGE DATA UN SMP SEKOTA SEMARANG……35

GAMBAR 4.12 DENDOGRAM COMPLETE LINKAGE UN SEKOTA SEMARANG……36

GAMBAR 4.13 CLUSTERING AVERAGE LINKAGE DATA UN SEKOTA SEMARANG……37

GAMBAR 4.14 DENDOGRAM AVERAGE LINKAGE UN SEKOTA SEMARANG……37

GAMBAR 4.15 RINGKASAN NILAI UN SMP SWASTA……38

GAMBAR 4.16 CLUSTERING COMPLETE LINKAGE DATA UN SMP SWASTA……...39

GAMBAR 4.17 DENDOGRAM COMPLETE LINKAGE NILAI UN SMP SWASTA……...39

GAMBAR 4.18 CLUSTERING AVERAGE LINKAGE NILAI UN SMP SWASTA……...40

GAMBAR 4.19 DENDOGRAM AVERAGE LINKAGE UN SMP SWASTA...41

GAMBAR 4.20 RINGKASAN NILAI UN SMP NEGERI……..42

GAMBAR 4.21 CLUSTERING COMPLETE LINKAGE NILAI UN SMP NEGERI…….42

GAMBAR 4.22 DENDOGRAM COMPLETE LINKAGE DATA UN SMP NEGERI…….43

GAMBAR 4.23 CLUSTERING AVERAGE LINKAGE NILAI UN SMP NEGERI…….44

GAMBAR 4.24 DENDOGRAM AVERAGE LINKAGE NILAI UN SMP NEGERI…….44

GAMBAR 4.25 RINGAKASAN DATA UN BAHASA INDONESIA……45

GAMBAR 4.26 CLUSTERING COMPLETE LINKAGE UN BAHASA INDONESIA……..45

GAMBAR 4.27 DENDOGRAM COMPLETE LINKAGE UN BAHASA INDONESIA……..46

GAMBAR 4.28 CLUSTERING AVERAGE LINKAGE UN BAHASA INDONESIA……..47

GAMBAR 4.29 DENDOGRAM AVERAGE LINKAGE UN BAHASA INDONESIA……..47

GAMBAR 4.30 RINGKASAN NILAI UN BAHASA INGGRIS……48

GAMBAR 4.31 CLUSTERING COMPLETE LINKAGE NILAI UN BAHASA INGGRIS……48

GAMBAR 4.32 DENDOGRAM COMPLETE LINKAGE UN BAHASA INGGRIS……49

GAMBAR 4.33 CLUSTERING AVERAGE LINKAGE NILAI UN BAHASA INGGRIS……50

GAMBAR 4.34 DENDOGRAM AVERAGE LINKAGE UN BAHASA INGGRIS……50

GAMBAR 4.35 RINGKASAN NILAI UN MATEMATIKA……...51

GAMBAR 4.36 CLUSTERING COMPLETE LINKAGE NILAI UN MATEMATIKA……51

GAMBAR 4.37 DENDOGRAM COMPLETE LINKAGE NILAI UN MATEMATIKA……52

GAMBAR 4.38 CLUSTERING AVERAGE LINKAGE NILAI UN MATEMATIKA……53

GAMBAR 4.39 DENDOGRAM AVERAGE LINKAGE NILAI UN MATEMATIKA……53

GAMBAR 4.40 RINGKASAN NILAI UN IPA……54

GAMBAR 4.41 CLUSTERING COMPLETE LINKAGE NILAI UN IPA…….54

GAMBAR 4.42 DENDOGRAM COMPLETE LINKAGE NILAI UN IPA……55

GAMBAR 4.43 CLUSTERING AVERAGE LINKAGE NILAI UN IPA……...55

GAMBAR 4.44 DENDOGRAM AVERAGE LINKAGE NILAI UN IPA……..56

DAFTAR RUMUS

RUMUS 2.1 AVERAGE LINKAGE……...……..14

RUMUS 2.2 COMPLETE LINKAGE……...……14

RUMUS 2.3 EUCLIDEAN DISTANCE…….……..16

DAFTAR LAMPIRAN

Lampiran 1 Lembar Bimbingan 1

Lampiran 2 Lembar bimbingan 2

Lampiran 3 Hasil turnitin abstrak

Lampiran 4 Hasil turnitin Bab I

Lampiran 5 Hasil turnitin naskah publikasi

Lampiran 6 Hasil turnitin Skripsi

Lampiran 7 Surat bebas plagiasi