



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**Pemetaan Level Kemampuan Bahasa Inggris  
Kota-Kota di Indonesia dengan Teknik Cluster**

**SKRIPSI**



**Diajukan Guna Memenuhi Persyaratan Memperoleh  
Gelar Sarjana Komputer (S.Kom.)**

**Fededi**

**11110310029**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2015**

## PERSETUJUAN LAPORAN SKRIPSI

### **Pemetaan Level Kemampuan Bahasa Inggris Kota-Kota di Indonesia dengan Teknik Cluster**

Oleh

Nama : Fededi

NIM : 11110310029

Fakultas : Teknologi Informasi dan Komunikasi

Program Studi : Sistem Informasi

Tangerang, 10 Juli 2015

**Pembimbing**

Ir. Raymond Sunardi Oetama, MCIS.

Mengetahui,

**Ketua Program Studi Sistem Informasi**

Wira Munggana, S.Si.,M.Sc.

# HALAMAN PENGESAHAN

Skripsi dengan judul

## **Pemetaan Level Kemampuan Bahasa Inggris Kota-Kota di Indonesia dengan Teknik Cluster**

Oleh  
Fededi

Telah diujikan pada hari Jumat, tanggal 7 Agustus 2015, Pukul 13.00 s.d 14.30  
dan dinyatakan lulus dengan susunan penguji sebagai berikut

**Ketua Sidang**

**Penguji**

Yustinus Eko Soelistio, S.Kom., M.M.

Wira Munggana, S.Si., M.Sc.

**Pembimbing**

Ir. Raymond Sunardi Oetama, MCIS.

Disahkan oleh,

**Ketua Program Studi Sistem Informasi  
Universitas Multimedia Nusantara**

Wira Munggana, S.Si., M.Sc.

## **PERNYATAAN TIDAK MELAKUKAN PLAGIAT**

Dengan ini saya:

Nama : Fededi

NIM : 11110310029

Program Studi : Sistem Informasi

Dengan ini saya menyatakan bahwa skripsi ini adalah karya ilmiah saya sendiri, dan saya tidak melakukan plagiat. Semua kutipan karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah saya sebutkan sumber kutipannya serta saya cantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan atau penyimpangan baik dalam pelaksanaan maupun dalam penulisan skripsi, saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS** untuk mata kuliah skripsi yang telah saya tempuh.

Tangerang, 10 Juli 2015

Fededi

UMMN

# Pemetaan Level Kemampuan Bahasa Inggris Kota-Kota di Indonesia dengan Teknik Cluster

## 11 ABSTRAKSI

Bahasa Inggris merupakan bahasa yang menjadi fokus Universitas Multimedia Nusantara (UMN) untuk dikembangkan dalam mempersiapkan mahasiswanya menghadapi Masyarakat Ekonomi Asean (MEA) pada bulan desember 2015 sehingga lulusannya dapat bersaing dengan negara lain. Untuk mendukung fokus tersebut maka diadakan penelitian ini untuk menguji serta melakukan pemetaan level kemampuan berbahasa Inggris mahasiswa UMN. Teknik cluster merupakan teknik yang ada didalam data mining digunakan untuk mengelompokkan data yang memiliki kesamaan sehingga teknik cluster ini dipakai untuk pemetaan level kemampuan bahasa inggris kota-kota di Indonesia dengan menggunakan algoritma K-means. Tujuan dari penelitian ini menghasilkan cluster kemampuan bahasa Inggris dan analisis kesenjangan kemampuan bahasa Inggris di kota-kota madya/kabupaten di Indonesia. Metodologi yang digunakan adalah *Knowledge Discovery of Data* (KDD) karena lebih lengkap dan sesuai dengan penelitian serta proses pengerjaan ini akan dibantu oleh aplikasi WEKA yang akan membentuk cluster – cluster sesuai kebutuhan. Hasil dari *cluster* adalah pemetaan kota-kota Indonesia yang memiliki kesamaan dan disimpan dalam format *microsoft excel* untuk diberikan ke proses berikutnya data *visualization*.

Kata kunci: Bahasa Inggris, UMN, teknik cluster, K-means, *Knowledge Discovery of Data* , WEKA

# **The Mapping of English Ability Level Amongst Cities in Indonesia using Cluster Technique**

## **13 ABSTRACT**

English is the language that becomes focus for Multimedia Nusantara University (UMN) to develop its students in preparation to face the Asean Economic Community (AEC) in December 2015. So that UMN graduates are able to compete with other jobseekers from other countries. To support this focus this study is conducted to examine and map the levels of English language proficiency of UMN students. Cluster technique is a technique in data mining that is used to process the data. Clusters are build using the K-means algorithm to map cities with similar levels in English language ability. The clusters also give other information such as English language skills gaps amongst cities in Indonesia. The methodology used is Knowledge Discovery of Data (KDD) on Weka application.

Keywords: English, UMN, Clusters technique, K-means, Knowledge Discovery of Data, WEKA

# UMN

## KATA PENGANTAR

Riset kemampuan bahasa Inggris mahasiswa Universitas Multimedia Nusantara dan laporan skripsi ini tidak akan bisa penulis selesaikan tepat pada waktunya tanpa ada bimbingan dari Sang Juru Selamat. Laporan skripsi yang berjudul “Pemetaan Level Kemampuan Bahasa Inggris Kota-Kota Di Indonesia dengan Teknik Cluster” diajukan kepada Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Multimedia Nusantara.

Laporan skripsi ini juga tidak akan bisa selesai tanpa adanya bantuan dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Wira Mungana, S. Si., M.Sc. selaku Ketua Program Studi Sistem Informasi di Universitas Multimedia Nusantara.
2. Ir. Raymond Sunardi Oetama, MCIS. selaku Dosen Pembimbing yang selalu memberikan banyak dukungan dan inspirasi dalam pembuatan riset dan penulisan laporan skripsi.
3. Bapak Feris Thia dan Ibu Yovita Surianto yang memberikan ide awal untuk proses riset ini.
4. Orang tua, saudara, dan keluarga yang selalu memberikan dukungan semangat dan doa selama mengerjakan laporan skripsi
5. Meriani dan keluarga yang selalu memberikan dukungan dan doa selama mengerjakan laporan skripsi.



6. Yuri Pramana, Wisnu Satyagraha, dan Theodora Giovani selaku sahabat penulis dan rekan riset yang telah melewati suka duka dalam mengerjakan laporan skripsi..
7. Ryan Susanto, Patriazka Odie, Kevin Adhinegoro, dan Candy Claudia yang bersedia untuk sharing dan cerita selama pembuatan riset dan penulisan laporan skripsi.
8. Ivan Dermawan, Dionisius Susanto, Daniel Nalendra, Aril Rulif, dan teman – teman lain di Wisma Disuka 16 yang bersedia untuk berbagi cerita tentang pengalaman mengerjakan laporan skripsi..

Semoga isi dari laporan skripsi ini dapat memberikan wawasan dan inspirasi bagi pembaca khususnya bagi mahasiswa UMN.

Tangerang, 10 Juli 2015

Penulis

UMMN

# DAFTAR ISI

PERSETUJUAN LAPORAN SKRIPSI .....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN TIDAK MELAKUKAN PLAGIAT .....	iv
ABSTRAKSI .....	v
ABSTRACT.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	7
2.1 Knowledge Discovery of Data .....	7
2.2 Data Mining.....	10
2.3 Pengelompokan Data Mining.....	11
2.4 Clustering .....	12
2.4.1 K-means Clustering.....	12
2.5 Data Preprocessing .....	13
2.6 Perbandingan Proses Model Data Mining.....	14
2.7 Perbandingan Aplikasi yang dipakai pada Data Mining .....	16
2.8 Data Masukan.....	16

2.8.1	ARFF.....	17
2.8.2	CSV.....	18
2.8.3	UCI.....	18
2.8.4	XLS.....	18
2.9	Pendidikan.....	19
2.10	Bahasa Inggris.....	19
2.11	Manfaat Bahasa Inggris.....	20
2.12	TOEFL.....	21
2.13	WEKA.....	21
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>23</b>
3.1	Gambaran Umum Objek Penelitian.....	23
3.2	Visi dan Misi Univesitas Multimedia Nusantara.....	24
3.3	Stuktur Organisasi.....	25
3.4	Kebutuhan riset tiap tahapan.....	26
3.5	Metode Penelitian.....	27
3.6	Atribute Penelitian.....	29
3.7	Teknik Pengumpulan Data.....	31
3.8	Teknik Pengambilan Sampel.....	32
3.9	Teknik Analisa Data.....	32
<b>BAB IV ANALISIS DAN HASIL PENELITIAN.....</b>		<b>33</b>
4.1	Data dari Data <i>Warehouse</i> .....	33
4.2	Selection dan Transformation data.....	33
4.3	Proses Data Mining dengan Teknik <i>Cluster</i> .....	40
4.3.1	Langkah – langkah memakai WEKA.....	40
4.3.2	Proses Clustering dengan tabel rata-rata tidak <i>Grouping</i> .....	45
4.3.3	Proses Clustering dengan tabel rata-rata <i>group</i> .....	49
4.3.4	Proses Clustering dengan tabel data keseluruhan.....	51
4.4	Evaluasi <i>Cluster</i> .....	51

4.5 Hasil Pemetaan Kota .....	53
4.6 Perbandingan pemetaan kota antara tabel` .....	60
BAB V KESIMPULAN.....	63
5.1 Kesimpulan.....	63
5.2 Saran.....	64
DAFTAR PUSTAKA .....	65
LAMPIRAN.....	67



## DAFTAR GAMBAR

Gambar 1.1 Level kemampuan Bahasa Inggris .....	2
Gambar 2.1 Proses <i>Knowledge Discovery of Data</i> (KDD).....	7
Gambar 2.2 Bidang Ilmu Data Mining .....	11
Gambar 2.3 Format dalam Arff.....	17
Gambar 2.4 Aplikasi Weka.....	12
Gambar 3.1 Struktur Organisasi Universitas Multimedia Nusantara.....	25
Gambar 3.2 Rancangan data model pada Data Warehouse .....	26
Gambar 3.2 Proses <i>Knowledge Discovery of Data</i> (KDD).....	27
Gambar 4.1 Tabel hasil kuis .....	34
Gambar 4.2 Tabel rata-rata <i>group</i> .....	35
Gambar 4.3 Tabel rata-rata tidak <i>group</i> .....	39
Gambar 4.4 Tabel data keseluruhan.....	39
Gambar 4.5 Tampilan Preprocess .....	40
Gambar 4.6 Menu Cluster.....	41
Gambar 4.7 Tampilan menu cluster.....	41
Gambar 4.8 Tampilan Algoritma pada cluster.....	42
Gambar 4.9 Tampilan Simplekmeans .....	42
Gambar 4.10 Cluster Output .....	43
Gambar 4.11 Visualize Cluster .....	43
Gambar 4.12 Tampilan Visualize Cluster.....	44
Gambar 4.13 Tampilan pilih edit .....	44
Gambar 4.14 Sum of Square Errors tabel rata-rata tidak <i>group</i> .....	47
Gambar 4.15 Total SSE cluster 4 seed 40.....	52
Gambar 4.16 Grafik tabel rata-rata <i>group</i> .....	56

Gambar 4.17 Penyebaran anggota data keseluruhan level 1.....	56
Gambar 4.18 Penyebaran anggota data keseluruhan level 2.....	57
Gambar 4.19 Penyebaran anggota data keseluruhan level 3.....	58
Gambar 4.20 Penyebaran anggota data keseluruhan level 4.....	60



## DAFTAR TABEL

Tabel 2.1 Summary KDD, CRISP-DM, dan SEMMA Proses.....	15
Tabel 3.1 Attribute .....	29
Tabel 4.1 Rata-rata <i>group</i> .....	35
Tabel 4.2 Rata-rata tidak <i>group</i> .....	36
Tabel 4.3 Total SSE .....	46
Tabel 4.4 Persentase dan jumlah anggota cluster dengan $k=4$ dan $s=40$ .....	47
Tabel 4.5 Jumlah Iterasi .....	47
Tabel 4.6 Total SSE .....	49
Tabel 4.7 Persentase dan jumlah anggota tabel rata-rata tidak <i>group</i> .....	50
Tabel 4.8 Hasil <i>Cluster</i> tabel rata-rata <i>group</i> .....	50
Tabel 4.9 Persentase dan jumlah anggota tabel data keseluruhan .....	51
Tabel 4.10 Data Keseluruhan level 1 .....	57
Tabel 4.11 Data Keseluruhan level 2 .....	58
Tabel 4.12 Data Keseluruhan level 3 .....	59
Tabel 4.13 Data Keseluruhan level 4 .....	59

UMMN