

**IMPLEMENTASI *AGGLOMERATIVE HIERARCHIAL CLUSTERING*
UNTUK *CLUSTERING* PERTANYAAN PENGGUNA
PADA RUMAH 2 IN 1**

SKRIPSI



disusun oleh

Hisyam Mustofa

14.11.8137

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2017**

**IMPLEMENTASI *AGGLOMERATIVE HIERARCHIAL CLUSTERING*
UNTUK *CLUSTERING* PERTANYAAN PENGGUNA
PADA RUMAH 2 IN 1**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Hisyam Mustofa

14.11.8137

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2017**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI *AGGLOMERATIVE HIERARCHIAL CLUSTERING*
UNTUK CLUSTERING PERTANYAAN PENGGUNA
PADA RUMAH 2 IN 1**

yang dipersiapkan dan disusun oleh

Haryam Mustofa

14.11.8137

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 28 September 2017

Dosen Pembimbing,



Yuli Astuti, M.Kom.

NIK. 190302146

PENGESAHAN
SKRIPSI
IMPLEMENTASI AGGLOMERATIVE HIERARCHIAL CLUSTERING
UNTUK CLUSTERING PERTANYAAN PENGGUNA
PADA RUMAH 2 IN 1

yang dipersiapkan dan disusun oleh

Hisyam Mustofa

14.11.8137

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 November 2017

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Sudarmawan, S.T., M.T.
NIK. 190302035



Bayu Setiaji, M.Kom.
NIK. 190302216



Barka Satya, M.Kom.
NIK. 190302126



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 November 2017

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di 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 tanggung jawab penulis pribadi.



MOTTO

"suatu hari di mana angin bergerak sempurna

suatu layar hanya perlu dibuka

dan dunia ini penuh dengan keindahan

hari ini adalah hari tersebut"

~ Mevlana Rumi, 13th Century



PERSEMBAHAN

Skripsi ini aku persembahkan kepada:

- ALLAH Subhanahu Wa Ta'ala
- Diriku sendiri dan keluarga
- Pihak Rumah 2 In 1
- Bu Yuli Astuti selaku dosbimku
- Semua teman yang pernah dan sedang berinteraksi denganku
- Ahli Spiritual dan Ahli Mistis
- Hujan di Bulan November yang selalu setia menemaniku
- Alam semesta yang lemah lembut

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah atas berkah dan rahmat Nya sehingga penulis dapat menyelesaikan Skripsi Berjudul “Implementasi *Agglomerative Hierarchial Clustering* Untuk *Clustering* Pertanyaan Pengguna Pada Rumah 2 In 1.”

Skripsi ini merupakan salah satu persyaratan kelulusan jenjang Program Sarjana Strata 1 pada jurusan Informatika pada Universitas Amikom Yogyakarta.

Dengan selesainya penyusunan skripsi ini, maka penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Sudarmawan, M.T. selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Ibu Yuli Astuti, M. Kom. selaku dosen pembimbing.
5. Pihak Rumah 2 in 1 selaku objek penelitian.
6. Keluarga, teman, dan segenap pihak yang telah membantu menyelesaikan skripsi ini.

Yogyakarta, 29 November 2017

Penulis,

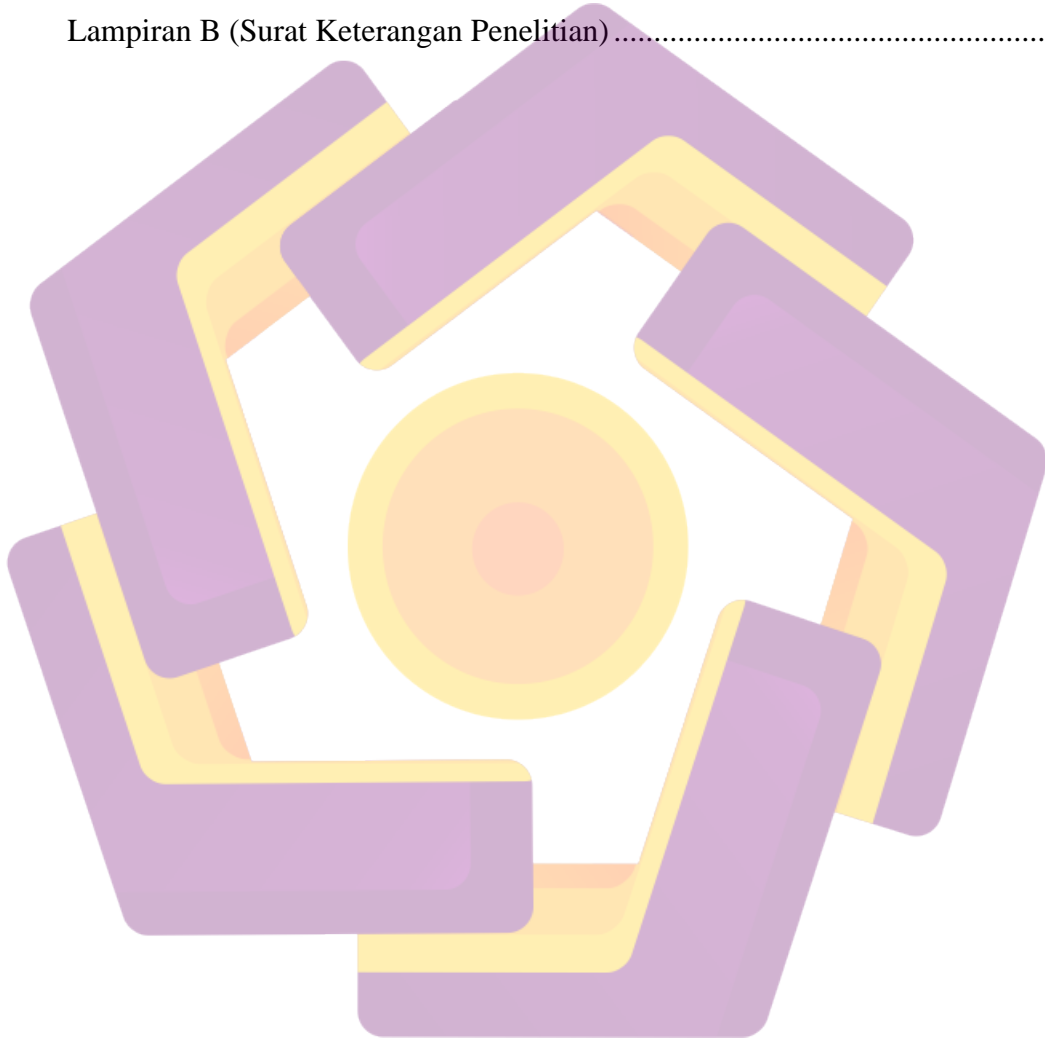
DAFTAR ISI

JUDUL.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
INTISARI.....	xiv
ABSTRACT	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Maksud dan Tujuan Penelitian.....	5
1.5 Metode Penelitian.....	5
1.5.1 Metode Pengumpulan Data	5
1.5.2 Metode Analisis Data.....	6
1.5.3 Metode Pengembangan Sistem.....	7
1.5.4 Metode Pengujian	7
1.6 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI.....	11
2.1 Tinjauan Pustaka	11

2.2	Dasar Teori	15
2.2.1	Data Mining.....	15
2.2.2	Text Mining.....	18
2.2.3	Analisis <i>Cluster</i>	21
2.2.4	Algoritma Agglomerative Hierarchial <i>Clustering</i>	24
2.2.5	Information Retrieval	28
2.2.6	Metode Stemming	29
2.2.7	Metode Porter Stemmer (Bahasa Indonesia)	29
2.2.8	Fungsi Stemming.....	33
2.2.9	Pengertian Tokenizing	34
2.2.10	Stopwords, Stoplist, Stopwords Removal	35
2.2.11	Term Frequency – Inverse Document Frequency (tf-idf)	39
2.2.12	Koefisien Korelasi <i>Cophenetic</i>	40
BAB III ANALISIS DAN PERANCANGAN		41
3.1	Tinjauan Umum.....	41
3.1.1	Profile Singkat Usaha.....	41
3.1.2	Visi dan Misi Usaha	41
3.1.3	Identifikasi Penyebab Masalah	42
3.2	Analisis Kebutuhan Sistem	43
3.2.1	Kebutuhan Fungsional	43
3.2.2	Kebutuhan Non Fungsional.....	44
3.3	Analisis Data	46
3.4	Perancangan Sistem.....	54
3.4.1	Diagram <i>Use Case</i>	54
3.4.2	Diagram <i>Activity</i>	55
3.4.3	Diagram <i>Class</i>	62
3.4.4	Diagram <i>Sequence</i>	63
3.4.5	Perancangan Tabel.....	68
3.5	Perancangan Antarmuka Pengguna.....	72
3.5.1	Antarmuka Login.....	72

3.5.2	Antarmuka Menu Admin.....	73
3.5.3	Antarmuka Menu Data Analisis	73
3.5.4	Antarmuka Home (Dashboard).....	74
3.5.5	Antarmuka Analisis.....	75
3.5.6	Antarmuka Hasil Analisis (Kelola).....	77
3.5.7	Antarmuka Kelola User.....	78
BAB IV IMPLEMENTASI DAN PEMBAHASAN		80
4.1	Basis Data.....	80
4.1.1	Tabel Role.....	80
4.1.2	Tabel User.....	80
4.1.3	Tabel Vectorspace.....	81
4.1.4	Tabel Klustermodel.....	81
4.1.5	Tabel Kelompok.....	82
4.1.6	Tabel Cleandata.....	82
4.2	Fungsi-Fungsi Kode.....	83
4.2.1	Koneksi Database.....	83
4.2.2	Kode Autentikasi.....	84
4.2.3	Kode Main.....	84
4.2.4	Kode Analisis.....	85
4.3	Halaman Web (Interface).....	86
4.3.1	Halaman Login.....	86
4.3.2	Halaman Analisis.....	86
4.3.3	Halaman Home (Dashboard).....	90
4.3.4	Halaman Hasil <i>Cluster</i>	93
4.3.5	Halaman Kelola.....	94
4.4	Evaluasi Model.....	96
4.5	Pengujian.....	98
4.5.1	White Box Testing.....	98
4.5.2	Black Box Testing.....	99
BAB V PENUTUP.....		102

5.1 Kesimpulan.....	102
5.2 Saran.....	103
DAFTAR PUSTAKA	105
LAMPIRAN.....	1
Lampiran A (Daftar Pertanyaan).....	1
Lampiran B (Surat Keterangan Penelitian).....	4



DAFTAR TABEL

Tabel 2.1	Hasil perbandingan Precision dan Recall	12
Tabel 2.2	Tinjauan Pustaka	13
Tabel 3.1	Tabel Performance Requirements	44
Tabel 3.2	Tabel spesifikasi minimal hardware	45
Tabel 3.3	Tabel spesifikasi minimal software	45
Tabel 3.4	Tabel spesifikasi minimal brainware	46
Tabel 3.5	Persiapan Data	47
Tabel 3.6	Proses Tokenisasi	47
Tabel 3.7	Data tanpa stopwords	48
Tabel 3.8	Data diubah kebentuk dasar	48
Tabel 3.9	Vektorisasi data	50
Tabel 3.10	Jarak antar dokumen	51
Tabel 3.11	Struktur tabel Role	69
Tabel 3.12	Struktur tabel User	69
Tabel 3.13	Struktur tabel Vectorspace	70
Tabel 3.14	Struktur tabel Klustermodel	70
Tabel 3.15	Struktur tabel Kelompok	71
Tabel 3.16	Struktur tabel Cleandata	72
Tabel 4.1	Hasil dari Whitebox Testing (Unit testing)	99
Tabel 4.2	Hasil dari Blackbox Testing	100

DAFTAR GAMBAR

Gambar 1.1	Alur pengembangan sistem menggunakan metode Waterfall	7
Gambar 2.1	Proses Preprocessing Text Mining	19
Gambar 2.2	Dendogram model pada Hierarchial Clustering	22
Gambar 2.3	Tahapan pada proses Document (text) Clustering.....	24
Gambar 2.4	Pseudocode Algoritma Agglomerative.....	25
Gambar 2.5	Algoritma porter stemmer untuk Bahasa Indonesia	30
Gambar 2.6	Contoh proses tokenizing	34
Gambar 2.7	Tahapan proses stopwords	36
Gambar 2.8	Hasil proses stopwords.....	37
Gambar 2.9	Tahapan pada proses stopwords removal.....	38
Gambar 2.10	Hasil proses stopwords removal	38
Gambar 3.1	Clustering Tahap 1	52
Gambar 3.2	Clustering Tahap 2	53
Gambar 3.3	Diagram Use Case	54
Gambar 3.4	Diagram Authentikasi Activity.....	55
Gambar 3.5	Diagram Analisis Hirarkis Activity.....	56
Gambar 3.6	Diagram Analisis Cluster Activity	57
Gambar 3.7	Diagram Kelola Activity	58
Gambar 3.8	Diagram Kelola Tambah User Activity.....	59
Gambar 3.9	Diagram Kelola Delete User Activity	60
Gambar 3.10	Diagram Kelola Activity	61
Gambar 3.11	Diagram Class	62
Gambar 3.12	Diagram Sequence proses Authentikasi	63
Gambar 3.13	Diagram Sequence proses Analisis Hirarkis	64
Gambar 3.14	Diagram Sequence proses Analisis Cluster.....	65
Gambar 3.15	Diagram Sequence proses Hasil	66
Gambar 3.16	Diagram Sequence proses Kelola (admin)	67

Gambar 3.17	Diagram Sequence proses Kelola (admin)	68
Gambar 3.18	User Interface form login	72
Gambar 3.19	User Interface Menu Admin	73
Gambar 3.20	User Interface Menu Data Analis	74
Gambar 3.21	User Interface Dashboard	75
Gambar 3.22	User Interface hirarki cluster	76
Gambar 3.23	User Interface analisis cluster	77
Gambar 3.24	User Interface Hasil cluster	78
Gambar 3.25	User Interface list user	78
Gambar 3.26	User Interface tambah user	79
Gambar 4.1	Hasil Tabel Role pada Database	80
Gambar 4.2	Hasil Tabel User pada Database	81
Gambar 4.3	Hasil Tabel Vectorspace pada Database	81
Gambar 4.4	Hasil Tabel Klustermodel pada Database	82
Gambar 4.5	Hasil Tabel Kelompok pada Database	82
Gambar 4.6	Hasil Tabel Cleandata pada Database	83
Gambar 4.7	Kode koneksi ke database	83
Gambar 4.8	Kode pada Proses Authentifikasi	84
Gambar 4.9	Kode pada Proses Main	85
Gambar 4.10	Kode pada Proses Analisis	85
Gambar 4.11	Halaman Login	86
Gambar 4.12	Halaman Analisis Hirarkis	87
Gambar 4.13	Halaman Hasil Visualisasi Data dan Analisis Hirarkis	88
Gambar 4.14	Halaman Hasil Analisis Hirarki	89
Gambar 4.15	Halaman Analisis Cluster	90
Gambar 4.16	Halaman Dashboad (tidak aktif)	91
Gambar 4.17	Halaman Dashboard (aktif)	92
Gambar 4.18	Halaman Tabel Pengelompokkan	93
Gambar 4.19	Halaman Hasil	94
Gambar 4.20	Halaman List User	95
Gambar 4.21	Halaman Tambah User	95

Gambar 4.22 Evaluasi kecepatan proses 96
Gambar 4.23 Evaluasi koefisien korelasi copenetic..... 97
Gambar 4.24 Contoh kode unit testing..... 98



INTISARI

Hampir semua data yang ada di dalam suatu layanan berbasis *web* berbentuk teks. Dalam layanan berbasis *web*, data teks berfungsi sebagai pembentuk kata dan bahasa yang digunakan dalam komunikasi antara pengguna dan pemberi layanan. Rumah 2 in 1 adalah sebuah layanan berbasis *web* yang bergerak dibidang *ecommerce property*. Sebagai sebuah bisnis yang masih tergolong *startup*, Rumah 2 in 1 mempunyai kendala dalam masalah kuantitas SDM yang kurang memadai, sehingga banyak data, terutama data yang berbentuk teks tidak terorganisasi dengan baik. Sebagai contoh adalah data pertanyaan yang diajukan pengguna layanan ini. Penataan data pertanyaan dapat memberikan manfaat dalam membantu distribusi data ke pihak penjawab, dan memberi kualitas pelayanan yang lebih baik kepada pengguna dengan pembuatan FAQ.

Pada skripsi ini, peneliti mencoba untuk menganalisa peluang yang bisa diperoleh dengan mengorganisasi data teks, khususnya data pertanyaan dengan melakukan interview kepada pihak Rumah 2 in 1. Peneliti menggunakan metode *Tf-Idf* untuk memperoleh fitur pada data teks, *Cosine Similarity* sebagai penghitung jarak dan kemiripan antar dokumen (data teks), dan Algoritma *Agglomerative Hierarchical Clustering* untuk mengelompokkan data tersebut ke dalam *cluster* secara hirarkis.

Model *cluster* hirarkis yang dihasilkan mampu mengelompokkan data pertanyaan (data teks pendek) ke dalam kelompok-kelompok berdasarkan kemiripan teks dalam dokumen (yang diukur berdasarkan jarak antar dokumen). Untuk model *cluster* hirarkis tanpa proses *stemming*, mempunyai nilai korelasi *cophenetic* sebesar 0.6 untuk *single linkage*, 0.78 untuk *complete linkage*, dan 0.4 untuk *ward linkage*. Sedangkan model *cluster* hirarkis yang melalui proses *stemming*, mempunyai nilai korelasi *cophenetic* sebesar 0.62 untuk *single linkage*, 0.79 untuk *complete linkage*, dan 0.43 untuk *ward linkage*.

Kata-kunci: Analisis *Cluster*, *Agglomerative Hierarchical Clustering*, perancangan, *cosine similarity*, Sistem Temu Kembali Informasi, *Tf-Idf*,

ABSTRACT

Almost all the data contained in a web service is text-shaped. In web-based services, text data serves as the shaping of words and languages used in communication between users and service providers. Rumah 2 in 1 is a web-based service that is engaged in ecommerce property. As a business that is still a startup, Rumah 2 in 1 has constraints in the lack of human resources quantity problem, so much data, especially data in the form of text is not well organized. An example is the question data this service user asks. Arrangement of question data can provide benefits in assisting the distribution of data to the answering party, and provide better service quality to users by making FAQ.

In this research, researchers try to analyze the opportunities that can be obtained by organizing text data, especially data questions with interviews to the Rumah 2 in 1. Researchers use the method Tf-Idf to meperoleh feature on text data, Cosine Similarity as a counter distance and similarity between documents (text data), and Agglomerative Hierarchical Clustering algorithms to group the data into clusters in a hierarchical manner.

The resulting hierarchical cluster model is able to group question data (short text data) into groups based on the text similarity in the document (as measured by the distance between the documents). For hierarchical cluster model without stemming process, it has a cophenetic correlation value of 0.6 for single linkage, 0.78 for complete linkage, and 0.4 for ward linkage. While the hierarchical cluster model through the stemming process, has a value of 0.62 cophenetic correlation for single linkage, 0.79 for complete linkage, and 0.43 for ward linkage.

Keywords: *Cluster Analysis, Agglomerative Hirarchical Clustering, design, cosine similarity, information retrieval, Tf-Idf,*