

**PENGELOMPOKAN PROFIL PEKERJAAN ALUMNI MENGGUNAKAN
ALGORITMA K-MEANS**

SKRIPSI

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam



Oleh :

RATNA WIDIYASARI

1404730

PROGRAM STUDI ILMU KOMPUTER

DEPARTEMEN PENDIDIKAN ILMU KOMPUTER

FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS PENDIDIKAN INDONESIA

2019

**PENGELOMPOKAN PROFIL PEKERJAAN ALUMNI MENGGUNAKAN
ALGORITMA K-MEANS**

Oleh :

Ratna Widiyasaki

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Ratna Widiyasaki

Universitas Pendidikan Indonesia

Juni 2019

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis.

RATNA WIDIYASARI

1404730

**PENGELOMPOKAN PROFIL PEKERJAAN ALUMNI MENGGUNAKAN
ALGORITMA K-MEANS**

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Prof. Dr. H. Munir, M.IT.

NIP. 196603252001121001

Pembimbing II,

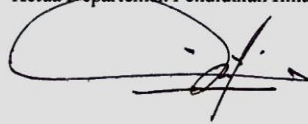


Dr. Rani Megasari, M.T.

NIP. 198705242014042002

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



Dr. Lala Septem R, M.T.

NIP. 197809262008121001

PERNYATAAN

Saya menyatakan bahwa skripsi dengan judul “Pengelompokan Profil Pekerjaan Alumni Menggunakan Algoritma K-Means” ini beserta seluruh isinya adalah hasil karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan dan tidak ada bagian skripsi ini yang merupakan tindakan plagiarisme. Atas pernyataan ini, saya siap menanggung risiko atau sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Desember 2019
Yang Membuat Pernyataan,

Ratna Widiyarsi
1404730

PENGELOMPOKAN PROFIL PEKERJAAN ALUMNI MENGGUNAKAN ALGORITMA K-MEANS

Oleh :

Ratna Widiyasari – widiyasari@student.upi.edu

1404730

ABSTRAK

Tracer Study adalah salah satu pelacakan jejak kepada alumni yang umum dilakukan program studi di perguruan tinggi sebagai upaya dalam memperbaiki kualitas penyelenggaraan pendidikan. Terdapat beberapa kuesioner yang ditujukan kepada alumni, namun tanggapan sebagai umpan balik yang diberikan alumni masih terbilang cukup rendah. Penelitian ini bertujuan mengoptimalkan program *tracer study* yang dilakukan dengan cara mengelompokkan profil pekerjaan alumni agar dapat disesuaikan dengan kebutuhan penyebaran kuesioner. Metode yang digunakan dalam pengelompokkan profil pekerjaan alumni adalah *clustering* yang dalam penelitian ini menggunakan algoritma K-Means. Hasil dari penelitian ini adalah *cluster-cluster* profil pekerjaan alumni yang setiap anggota dalam *cluster* yang sama memiliki kriteria pekerjaan yang mirip.

Katakunci: *Tracer Study, Clustering, K-Means, Hamming Distance*

CLUSTERING OF ALUMNUS' WORK PROFILES USING THE K-MEANS ALGORITHM

Arranged by

Ratna Widiyasari – widiyasari@student.upi.edu

1404730

ABSTRACT

Tracer Study is one of methods used in university to track their alumnus' traces as an approach to improve the quality of their education management. There exist a few questionnaires aimed at the alumnus, but responses the alumnus given are still quite lacking. This research focused on optimizing tracer study program by separating alumnus' work profiles into parts so it could suit distribution of the questionnaire. Method used to group the alumnus work profiles is clustering with the help of K Means algorithm. The aforementioned research resulting in clusters of alumnus' work profiles in which each member of the same cluster has similar work characteristics.

Keywords: *Tracer Study, Clustering, K-Means, Hamming Distance*

DAFTAR ISI

| | |
|--------------------------------------|------|
| ABSTRAK | ii |
| ABSTRACT | iii |
| KATA PENGANTAR | iv |
| UCAPAN TERIMAKASIH..... | v |
| DAFTAR ISI..... | viii |
| DAFTAR GAMBAR | x |
| DAFTAR TABEL..... | xi |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Tujuan Penelitian | 4 |
| 1.4 Manfaat Penelitian | 5 |
| 1.5 Batasan Masalah..... | 5 |
| 1.6 Sistematika Penulisan..... | 6 |
| BAB II KAJIAN PUSTAKA | 7 |
| 2.1 Data Mining | 7 |
| 2.2 Pengelompokkan (Clustering)..... | 10 |
| 2.3 Algoritma K-Means | 17 |
| 2.4 Hamming Distance..... | 21 |
| 2.5 Sum Squared Error (SSE) | 22 |
| 2.6 Elbow Method..... | 22 |
| BAB III METODOLOGI PENELITIAN..... | 25 |
| 3.1 Desain Penelitian..... | 25 |
| 3.2 Metode Penelitian..... | 25 |

| | | |
|-----------------------------------|---|----|
| 3.2.1 | Metode Pengumpulan Data | 27 |
| 3.2.2 | Metode Pengembangan Perangkat Lunak | 28 |
| 3.3 | Alat dan Bahan Penelitian | 30 |
| 3.3.1 | Alat Penelitian | 30 |
| 3.3.2 | Bahan Penelitian | 31 |
| BAB IV HASIL DAN PEMBAHASAN | | 32 |
| 4.1 | Pengumpulan Data | 32 |
| 4.2 | Perancangan Data Preparation | 33 |
| 4.3 | K-means Clustering | 36 |
| 4.4 | Pengembangan Perangkat Lunak | 37 |
| 4.4.1 | Analisa Kebutuhan | 37 |
| 4.4.2 | Desain | 38 |
| 4.4.3 | Parameter Analisis | 40 |
| 4.5 | Analisis Hasil Pengujian | 40 |
| 4.5.1 | Analisis Profil | 40 |
| 4.5.2 | Detail Profil | 43 |
| BAB V KESIMPULAN DAN SARAN | | 45 |
| 5.1 | Kesimpulan | 45 |
| 5.2 | Saran | 47 |
| DAFTAR PUSTAKA | | 50 |
| LAMPIRAN | | 50 |
| 1.1 | Lampiran 1. Tabel perhitungan jarak | 50 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Bidang Ilmu Data Mining | 7 |
| Gambar 2.2 Tahap-tahap pada Data Mining | 9 |
| Gambar 2.3 Tahapan Clustering | 11 |
| Gambar 2.4 Taxonomy Clustering | 12 |
| Gambar 2.5 Contoh Hierarchical Clustering..... | 13 |
| Gambar 2.6 Contoh Dendogram | 13 |
| Gambar 2.7 Contoh Dendogram Single, Complete, dan Average linkage | 14 |
| Gambar 4.1 Diagram alir data preparation..... | 33 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Teknik Berbasis Pengelompokkan Parsial..... | 15 |
| Tabel 4. 1 Perhitungan dengan hamming distance..... | 34 |
| Tabel 4. 2 Hasil Perhitungan dan Perbandingan Jaccard dan Hamming distance | 35 |

DAFTAR PUSTAKA

- Aina, L. O., & Moahi, K. (1999). Tracer study of the Botswana library school graduates. *Education for Information, 17*, 215–244.
- Alswaitti, M., Ishak, M. K., Ashidi, N., & Isa, M. (2018). Engineering Applications of Artificial Intelligence Optimized gravitational-based data *clustering* algorithm. *Engineering Applications of Artificial Intelligence, 73*(January 2017), 126–148. <https://doi.org/10.1016/j.engappai.2018.05.004>
- Aslan, B. (2009). Çeşitli Ülkelerde Yükseköğretimde Akreditasyon Uygulamaları ve Türk Yükseköğretimindeki Gelişmeler. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 42*(1), 287–309.
- Cao, F., Yu, L., Zhexue, J., & Liang, J. (2017). *k*-mw-modes : An algorithm for *clustering* categorical matrix-object data. *Applied Soft Computing Journal, 57*, 605–614. <https://doi.org/10.1016/j.asoc.2017.04.019>
- Cornuéjols, A., Wemmert, C., Gañçarski, P., & Bennani, Y. (2018). Collaborative *clustering*: Why, when, what and how. *Information Fusion, 39*, 81–95. <https://doi.org/10.1016/j.inffus.2017.04.008>
- Daniel, T. (n.d.). *An Introduction to Data Mining*.
- de França, F. O. (2016). A hash-based co-*clustering* algorithm for categorical data. *Expert Systems with Applications, 64*, 24–35. <https://doi.org/10.1016/j.eswa.2016.07.024>
- Elgamel, M. S., & Dandoush, A. (2015). A modified Manhattan distance with application for localization algorithms in ad-hoc WSNs. *Ad Hoc Networks, 33*, 168–189. <https://doi.org/10.1016/j.adhoc.2015.05.003>
- Fleming, S. J., Aspreem, E., & Taves, A. (2015). Published in, 4, 8–11.
- Gola, J., Britz, D., Staudt, T., Winter, M., Schneider, A. S., Ludovici, M., & Mücklich, F. (2018). Advanced microstructure classification by data mining methods. *Computational Materials Science, 148*(November 2017), 324–335.

<https://doi.org/10.1016/j.commatsci.2018.03.004>

Hou, J., & Liu, W. (2017). Parameter independent *clustering* based on dominant sets and *cluster* merging. *Information Sciences*, 405, 1–17.

<https://doi.org/10.1016/j.ins.2017.04.006>

Jenssen, R., & Eltoft, T. (2008). A new information theoretic analysis of sum-of-squared-error kernel *clustering*. *Neurocomputing*, 72(1–3), 23–31.

<https://doi.org/10.1016/j.neucom.2008.03.017>

Ji, J., Bai, T., Zhou, C., Ma, C., & Wang, Z. (2013). Neurocomputing An improved k-prototypes *clustering* algorithm for mixed numeric and categorical data.

Neurocomputing, 120, 590–596.

<https://doi.org/10.1016/j.neucom.2013.04.011>

Machine, P., & Tools, L. (n.d.). *No Title*.

Mesquita, D. P. P., Gomes, J. P. P., Souza Junior, A. H., & Nobre, J. S. (2017). Euclidean distance estimation in incomplete *datasets*. *Neurocomputing*, 248,

11–18. <https://doi.org/10.1016/j.neucom.2016.12.081>

Nasional, U. S. P. (2003). UU no. 20/2003 ttg Sisdiknas.

Negeri, K. L. (2015). Masyarakat Ekonomi ASEAN: Peluang & Tantangan, (November), 1–20.

Olson, C. F. (1995). Parallel algorithms for *hierarchical clustering*. *Parallel Computing*, 21(8), 1313–1325. [https://doi.org/10.1016/0167-8191\(95\)00017-I](https://doi.org/10.1016/0167-8191(95)00017-I)

I

Peffer, K. E. N., Tuunanen, T., & Rothenberger, M. A. (2008). A Design Science Research Methodology for Information Systems Research, 24(3), 45–77.

<https://doi.org/10.2753/MIS0742-1222240302>

Portugués, E. G. (2017). Lab notes for Statistics for Social Sciences II: Multivariate Techniques. Retrieved from <https://bookdown.org/egarpor/SSS2-UC3M/>

Setyaningsih, I. (2013). Tingkat Kepuasan Pengguna Lulusan. *Ilmiah Teknik*

Industri, 12(1), 73–82.

Shadroo, S., & Rahmani, A. M. (2018). US CR. *Computer Networks*.
<https://doi.org/10.1016/j.comnet.2018.04.001>

Surveys, A. C. M. C., & Dame, N. (1999). Data *clustering* : a review . ACM Comput
Surv Data *Clustering* : A Review, (March 2016).
<https://doi.org/10.1145/331499.331504>

Tom, G., & Elmer, L. (1994). Alumni Willingness to Give and Contribution
Behavior. *Journal of Services Marketing*, 8(2), 57–62.
<https://doi.org/10.1108/08876049410058442>

Zhu, E., & Ma, R. (2018). Ac ce pt us cr t. *Applied Soft Computing*.
<https://doi.org/10.1016/j.asoc.2018.07.026>

