PENGELOMPOKAN SKILL REQUIREMENT PADA LOWONGAN PEKERJAAN ALUMNI MENGGUNAKAN ALGORITMA K-MEANS

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer

oleh
Rizki Nugraha
1506748

PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG

2020

PENGELOMPOKAN *SKILL REQUIREMENT* PADA LOWONGAN PEKERJAAN ALUMNI MENGGUNAKAN ALGORITMA *K-MEANS*

Oleh
Rizki Nugraha
NIM 1506748

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

Universitas Pendidikan Indonesia

Januari 2020

RIZKI NUGRAHA

1506748

PENGELOMPOKAN *SKILL REQUIREMENT* PADA LOWONGAN PEKERJAAN ALUMNI MENGGUNAKAN ALGORITMA *K-MEANS*

Disetujui dan disahkan oleh pembimbing:

Pembimbing I,

Dr. Rani Megasari, M.T.

NIP. 198705242014042002

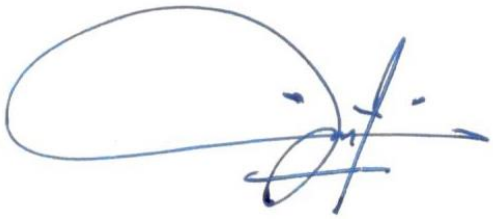
Pembimbing II,

Erna Piantari, S. Kom., M.T.

NIP. 920171219890224201

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



Lala Septem Riza, M.T., Ph. D.

NIP. 197809262008121001

PENGELOMPOKAN *SKILL REQUIREMENT* PADA LOWONGAN PEKERJAAN ALUMNI MENGGUNAKAN ALGORITMA *K-MEANS*

Oleh

Rizki Nugraha – nugraharizki80@gmail.com

1506748

ABSTRAK

Pengangguran merupakan salah satu permasalahan besar di Indonesia. Badan Pusat Statistik menunjukkan penurunan angka pengangguran di tingkat SD, SMP, SMK, dan SMA, namun justru mengalami peningkatan pada tingkat Diploma dan Sarjana. Salah satu faktor penyebab permasalahan ini adalah kesenjangan yang terjadi antara industri dengan pendidikan tinggi. Permasalahan ini dapat diatasi dengan memberikan lulusan pendidikan tinggi pengetahuan mengenai kebutuhan industri saat ini. Salah satunya dapat diperoleh melalui lowongan pekerjaan yang tersebar dalam sebuah media sosial, salah satunya grup alumni Departemen Pendidikan Ilmu Komputer UPI. Untuk mempermudah lulusan mendapatkan pengetahuan tersebut dilakukan pengelompokan lowongan pekerjaan dengan metode k-Means. Pemrosesan lowongan pekerjaan. Sebelum dapat dikelompokkan harus mengalami praproses terlebih dahulu. Praproses yang dilakukan menggunakan kerangka kerja text mining, langkah-langkahnya terdiri atas case folding, tokenizing, filtering, dan stemming. Khusus untuk stemming menggunakan algoritma Nazief dan Adriani karena menggunakan bahasa Indonesia. Hasil praproses direpresentasikan menjadi angka menggunakan algoritma TF-IDF dan ekstraksi fitur PCA untuk mereduksi dimensi. Setelah itu dilakukan pengelompokkan menggunakan algoritma k-Means dengan jumlah cluster yang bervariasi dan evaluasi menggunakan metode Elbow dan Silhouette Coefficient. Dari hasil analisis evaluasi cluster didapatkan nilai Sum of Squared Error sebesar 2.6517 dan nilai silhouette sebesar 0.5795 pada jumlah cluster 3.

Kata Kunci: *Clustering*, *Text Mining*, *TF-IDF*, lowongan pekerjaan, *Silhouette coefficient*, *Elbow method*

SKILL REQUIREMENT CLUSTERING IN ALUMNI JOB VACANCIES WITH K-MEANS ALGORITHM

Arranged by

Rizki Nugraha – nugraharizki80@gmail.com

1506748

ABSTRACT

Unemployment is one of the big problems in Indonesia. The Central Statistics Agency showed a decrease in unemployment at the elementary, junior high, vocational, and high school levels, but instead experienced an increase at the Diploma and Bachelor level. One factor causing this problem is the gap between the industry and higher education. This problem can be overcome by giving graduates of tertiary education knowledge of current industry needs. This knowledge can be obtained through job vacancies spread in a social media, one of them is the alumni group of UPI Computer Science Education Department. To make it easier for graduates to obtain this knowledge, a grouping of job vacancies is conducted using the k-Means method. Job vacancy processing Before it can be grouped, it must be preprocessed. The pre-processing is carried out using a text mining framework, the steps consist of case folding, tokenizing, filtering, and stemming. Especially for stemming using the Nazief and Adriani algorithm because it uses Indonesian. Preprocess results are represented as numbers using the TF-IDF algorithm and PCA feature extraction to reduce dimensions. After that the grouping is done using the k-Means algorithm with a variety of clusters and evaluation using the Elbow and Silhouette Coefficient methods. From the cluster evaluation analysis results obtained a Sum of Squared Error of 2.6517 and a silhouette value of 0.5795 on the number of clusters 3.

Keyword: *Clustering, Text Mining, TF-IDF, Job Vacancy, Silhouette Coefficient, Elbow Method*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR PUSTAKA

Abualigah, L. M., Khader, A. T., Al-Betar, M. A., & Alomari, O. A. (2017). Text Feature Selection With a Robust Weight Scheme and Dynamic Dimension Reduction to Text Document Clustering. *Expert Systems with Applications*, *84*, 24–36.

Afzali, M., & Kumar, S. (2019). Text Document Clustering: Issues and Challenges. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 263–268.

Agnihotri, D., Verma, K., & Tripathi, P. (2014). Pattern and Cluster Mining on Text Data. *2014 Fourth International Conference on Communication Systems and Network Technologies*, 428–432.

Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem Dan Informatika, 2009*, 196–201.

Beil, F., Ester, M., & Xu, X. (2002). Frequent Term-Based Text Clustering. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 436–442.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, *105*(9).

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.

Handayani, T. (2015). Relevansi lulusan perguruan tinggi di Indonesia dengan kebutuhan tenaga kerja di era global. *Jurnal Kependudukan Indonesia*, *10*(1), 53–64.

Hartono, O. R. (2016). Indonesian Stoplist. Retrieved January 12, 2019, from https://www.kaggle.com/oswinrh/indonesian-stoplist

Huang, A. (2008). Similarity Measures For Text Document Clustering. *Proceedings of The Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand*, *4*, 9–56.

Hudin, M. S., Fauzi, M. A., & Adinugroho, S. (2018). Implementasi Metode Text

Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(11), 5518–5524. Retrieved from http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3332

Iswahyuni, A. D. (2018). Desain Kurikulum Perguruan Tinggi untuk Mengeliminasi Gap Persepsi Perguruan Tinggi degan Industri. *Ratih: Jurnal Rekayasa Teknologi Industri Hijau*, *2*(2), 14.

Jolliffe, I. (2011). *Principal Component Analysis*. Springer.

Kusrini, K. (2015). Grouping of Retail Items by Using K-Means Clustering. *Procedia Computer Science*, *72*, 495–502. https://doi.org/10.1016/j.procs.2015.12.131

Langgeni, D. P., Baizal, Z. K. A., & AW, Y. F. (2015). Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection. *Seminar Nasional Informatika (SEMNASIF)*, *1*(4).

Liu, X.-W., He, P.-L., & Wang, H.-Y. (2005). The Research of Text Clustering Algorithms Based on Frequent Term Sets. *2005 International Conference on Machine Learning and Cybernetics*, *4*, 2352–2356.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

Mahmood, S., & Al-Rufaye, F. M. L. (2017). Arabic Text Mining Based on Clustering and Coreference Resolution. *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*, 140–144.

Nazief, B., & Adriani, M. (1996). Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. *Internal Publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta*, *41*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pusparisa, Y. (2019). Angka Pengangguran Lulusan Universitas Meningkat. Retrieved from https://katadata.co.id/infografik/2019/05/17/angka-

pengangguran-lulusan-perguruan-tinggi-meningkat

Robertson, S. (2004). Understanding Inverse Document Frequency: On Theoretical Arguments for IDF. *Journal of Documentation*, *60*(5), 503–520.

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Slamet, C., Rahman, A., Ramdhani, M. A., & Dharmalaksana, W. (2016). Clustering The Verses of The Holy Qur'an Using K-Means Algorithm. *Asian Journal of Information Technology*, *15*(24), 5159–5162. https://doi.org/10.3923/ajit.2016.5159.5162

Soleh, A. (2017). Masalah Ketenagakerjaan dan Pengangguran di Indonesia. *Cano Ekonomos*, *6*(2), 83–92.

Sommerville, I. (2010). *Software engineering*. New York: Addison-Wesley.

Statistik, B. P. (1999). *Berita Resmi Statistik*.

Statistik, B. P. (2019). *Berita Resmi Statistik*.

Sugianto, D. (2018). BI Catat Jumlah Lowongan Kerja di Kuartal II-2018 Naik 19,2%. Retrieved January 10, 2020, from https://finance.detik.com/berita-ekonomi-bisnis/d-4111902/bi-catat-jumlah-lowongan-kerja-di-kuartal-ii-2018-naik-192

Sukamto, R. A., & Shalahuddin, M. (2011). Rekayasa Perangkat Lunak Menggunakan CodeIgniter dan JQuery. *Yogyakarta: Andi*.

Thinsungnoen, T., Kaoungku, N., Durongdumronchai, P., Kerdprasop, K., & Kerdprasop, N. (2015). *The Clustering Validity with Silhouette and Sum of Squared Errors*. 44–51. https://doi.org/10.12792/iciae2015.012

Thomas, R. E., & Khan, S. S. (2016). Improved Clustering Technique Using Metadata for Text Mining. *2016 International Conference on Communication and Electronics Systems (ICCES)*, 1–5.

Tunali, V., & Bilgin, T. T. (2012). Examining the Impact of Stemming on Clustering Turkish Texts. *2012 International Symposium on Innovations in Intelligent Systems and Applications*, 1–4.

Vijayarani, S., Ilamathi, J., Nithya, M., Ilamathi, M. J., Nithya, M., Professor, A., & Research Scholar, M. P. (2015). Preprocessing Techniques for Text Mining

- An Overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16. https://doi.org/10.1016/j.procs.2013.05.286

Zhang, Y., & Jiang, M. (2010). Chinese Text Mining Based on Subspace Clustering. *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, *4*, 1617–1620.