

ANALISIS JENIS PEKERJAAN ALUMNI UNIVERSITAS MUHAMMADIYAH YOGYAKARTA MENGGUNAKAN ALGORITMA *K-MEANS*

(ANALYSIS OF THE TYPE OF ALUMNI WORK OF MUHAMMADIYAH UNIVERSITY OF
YOGYAKARTA USING *K-MEANS* ALGORITHM)

Astin Hasniaty, Slamet Riyadi, Asroni

ABSTRACT

Pendidikan harus berorientasi pada kompetensi yang dibutuhkan oleh dunia kerja karena persentase pengangguran di kalangan terdidik terus meningkat. Keberadaan alumni berperan penting dalam peningkatan kualitas yang telah dicapai oleh perguruan tinggi, meningkatnya lulusan pada Universitas Muhammadiyah Yogyakarta membuat tumpukan data semakin banyak, atas dasar masalah tersebut maka dilakukan pencarian pengetahuan baru dengan *data mining*. Pengelompokan data alumni akan dilakukan dengan metode *clustering* dan menggunakan algoritma *k-means*. Pada data alumni ini terdapat 5 atribut yang digunakan yaitu, prodi, IPK, tahun lulus, angkatan, dan jenis pekerjaan, di *cluster* untuk menemukan kelompok-kelompok yang memiliki kesamaan karakter, menganalisis program studi dengan setiap jenis pekerjaan. Hasil analisis ini menemukan kelompok jenis pekerjaan yang dapat dikatakan sangat mendekati dengan bidang studi. Analisis ini dilakukan dengan menggunakan *software RapidMiner Studio* dan sumber data diambil dari data alumni dalam bentuk data *set excel*. Analisis ini akan sangat membantu perkembangan mahasiswa yang akan lulus dan bekerja dengan memberikan strategi berupa persiapan memasuki dunia kerja.

Kata kunci: *data mining, k-means, clustering, alumni, RapidMiner Studio.*

PENDAHULUAN

Permasalahan di Indonesia yang saat ini sedang menjadi perhatian seluruh orang adalah pengangguran yang merajalela akibat fenomena perbedaan profesi seseorang yang tidak sesuai dengan pendidikan yang dipelajari. Persentase pengangguran di kalangan terdidik juga meningkat drastis. Pendidikan merupakan salah satu sasaran pokok pemerintah dalam rangka meningkatkan kesejahteraan rakyat. Pentingnya peranan perguruan tinggi untuk mengatasi masalah tersebut, perguruan tinggi yang berkualitas tentunya akan berusaha untuk mencetak lulusan yang siap bersaing menggunakan kemampuan dan kompetensi dalam dunia kerja. Keberadaan alumni berperan penting dalam peningkatan kualitas yang telah dicapai oleh perguruan tinggi, angka kelulusan sebuah universitas yang tinggi mempengaruhi data alumni perguruan tinggi tersebut.

Keberadaan alumni berperan penting dalam peningkatan kualitas yang telah dicapai oleh perguruan tinggi, karena semakin cepat lulusan mendapatkan pekerjaan itu artinya sistem belajar mengajar di universitas tersebut sudah cukup baik. Ditambah lagi jika pekerjaan alumni itu sesuai dengan jurusan semasa proses belajar. Tingginya angka kelulusan di Universitas Muhammadiyah Yogyakarta, maka data alumni semakin banyak juga yang masuk

dalam server database di Biro Sistem Informasi Universitas Muhammadiyah Yogyakarta. Oleh karena itu dalam pengelompokan data alumni Universitas Muhammadiyah Yogyakarta dibutuhkan penerapan data mining untuk menemukan informasi baru yang berguna bagi pihak universitas.

Berdasarkan uraian diatas, pengelompokan jenis pekerjaan alumni belum pernah dilakukan sebelumnya, oleh karena itu penulis akan menggunakan algoritma *k-means* dalam melakukan pengelompokan jenis pekerjaan alumni dengan hasil penyajian informasi baru pada data tersebut sesuai kebutuhan pihak Universitas Muhammadiyah Yogyakarta.

1. Rumusan Masalah

Berdasarkan latar belakang tersebut maka rumusan masalah penelitian ini adalah, *database* alumni dari Biro Sistem Informasi Universitas Muhammadiyah Yogyakarta masih berupa himpunan data dalam jumlah besar dan kompleks sehingga sulit untuk ditangani atau diproses jika hanya menggunakan manajemen basis data biasa atau aplikasi pemroses data tradisional. Perlu dilakukannya pengelompokan data yang mempunyai kemiripan karakter agar penyajian informasi pada data tersebut sesuai kebutuhan pihak Universitas Muhammadiyah Yogyakarta.

2. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah menganalisis pengelompokan jenis pekerjaan dengan program studi berdasarkan IPK, tahun lulus, dan angkatan alumni Universitas Muhammadiyah Yogyakarta menggunakan teknik *data mining* dengan algoritma *k-means*.

Data Mining

Data mining dalam aplikasinya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang memiliki tugas mengekstrak sebuah pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. Proses KDD sebagai berikut: (Angga Ginanjar Mabur 2012).

1. Data selection.

Pemilihan data / seleksi data dari sekumpulan data operasional perlu dilakukan sebelum masuk ke tahap penggalian informasi dalam KDD. Hasil data yang sudah diseleksi akan digunakan untuk proses *data mining*, dan disimpan dalam suatu berkas yang terpisah dari basis data operasional.

2. Pre-processing.

Ada tahapan sebelum masuk pada proses *data mining* yaitu proses *pre-processing* atau yang sering disebut *cleaning*. Yang bertujuan untuk membuang duplikasi data, memastikan data yang *inkonsisten*, dan memperbaiki kesalahan data, seperti kesalahan cetak / *tipografi*. *Enrichment* adalah proses “memperkaya” data yang sudah ada dengan informasi lain yang relevan dan di perlukan untuk KDD.

3. Transformation.

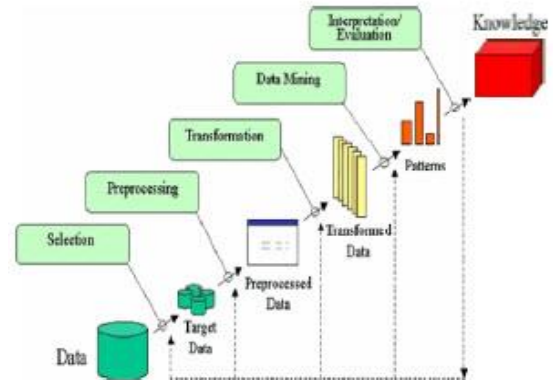
Proses *coding* pada data yang telah dipilih adalah transformasi. Data diproses hingga sesuai untuk *data mining*. Dalam KDD ada proses *coding* yang kreatif dan sangat tergantung pada jenis pola informasi yang akan dicari dalam basis data.

4. Data mining.

Data mining adalah sebuah proses mencari pola atau informasi menarik dalam data yang terpilih dengan menggunakan metode tertentu. Metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat tergantung tujuan dan proses KDD secara keseluruhan.

5. Interpretation / Evaluation.

Proses *data mining* yang dihasilkan oleh pola informasi perlu ditampilkan dalam bentuk yang mudah dimengerti. Tahap ini disebut *interpretation* yang merupakan bagian dari proses KDD. Pada tahap pemeriksaan ini mencakup pola atau informasi yang ditemukan apakah bertentangan dengan fakta atau hipotesis yang ada sebelumnya (Fina Nasari dan Surya Darma 2015).



Gambar 1. Tahapan data mining

Menurut (Linda Maulida 2018) *data mining* adalah suatu istilah yang digunakan untuk menemukan sebuah pengetahuan dalam *database* seperti pada gambar. 1 Proses *data mining* menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengeluarkan dan menganalisa informasi yang bermanfaat terkait pengetahuan dari berbagai *data base* besar.

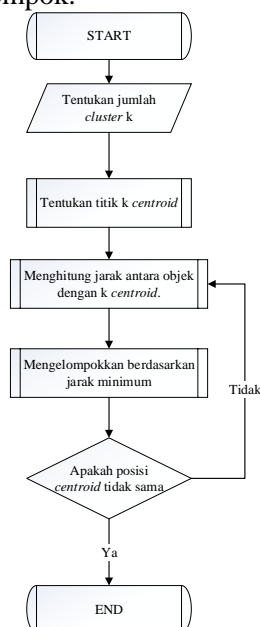
Clustering

Clustering adalah sesuatu metode untuk mencari dan mengelompokkan data yang memiliki persamaan karakteristik / *similarity* antara satu data dengan data yang lainnya. (Ong, Johan Oscar 2013) *Clustering* merupakan metode yang bersifat tanpa arah / *unsupervised*, maksudnya adalah metode ini diterapkan tanpa adanya latihan / *taining* dan tanpa ada guru / *teacher* serta tidak membutuhkan target *output*. Pengelompokan data-data kedalam beberapa jumlah kelompok / *cluster* berdasarkan struktur kelompok, keanggotaan data dalam kelompok, dan kekompakan data dalam kelompok. Ada dua jenis metode menurut pengelompokan, yaitu hierarki dan *partitioning*. Untuk hierarki, Satu data tunggal bisa dianggap sebuah kelompok, dua atau lebih. Sedangkan untuk *partitioning* membagi data menjadi satu kelompok. Jika menurut keanggotaan data, dibagi menjadi dua, yaitu eksklusif dan tumpang-tindih. Kategori eksklusif, data dipastikan hanya dapat menjadi satu kelompok.

Kategori tumpang-tindih data dapat menjadi anggota lebih dari satu kelompok. Berdasarkan kategori kekompakan ada komplet dan parsial. Data yang menyimpang disebut *outlier*, *noise*, atau *uninterested background*.

K-Means

K-means adalah metode pengelompokan data *non-hierarki* yang mempartisi data dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data yang berkarakter sama dan dikelompokkan menjadi satu dan kemudian data yang berkarakter beda akan dikelompokkan ke dalam kelompok lain. Tujuan pengelompokan ini adalah untuk meminimalkan fungsi objektif yang ada di *set* dalam suatu kelompok dan memaksimalkan variasi antar kelompok.



Gambar 2. Diagram Alir K-Means

Pengertian dari *k-means Clustering* adalah, *k* sebagai konstanta jumlah *cluster* yang diinginkan *Means* artinya nilai rata-rata dari suatu grup data yang didefinisikan sebagai *cluster*, sehingga *k-means clustering* adalah metode penganalisa *data mining* yang melakukan proses pemodelan tanpa supervisi / *unsupervised* dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi, seperti pada gambar 2. Metode *k-means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, data dalam satu kelompok mempunyai karakter yang sama satu sama lainnya dan data yang memiliki perbedaan karakteristik akan dipisahkan dalam kelompok yang berbeda. Dasar algoritma *k-means* adalah sebagai berikut:

1. Tentukan nilai *k* sebagai jumlah *cluster* yang akan dibentuk.
2. Inisialisasi *k* sebagai *centroid* yang dapat dibangkitkan secara *random*.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan persamaan *Euclidean Distance*

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Keterangan:

d = data titik dokumen (*euclidean*)

P = data *record*

Q = data *centroid*

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya*.
5. Tentukan posisi *centroid* baru (*k*).
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

RapidMiner Studio

RapidMiner Studio merupakan perangkat lunak yang bersifat terbuka (*open source*) yang memberikan solusi untuk melakukan analisis pada *data mining*, *text mining* dan analisis prediksi. *RapidMiner Studio* menggunakan berbagai macam cara deskriptif dan prediksi dalam memberikan pengetahuan pada pengguna sehingga dapat membuat keputusan yang terbaik. *RapidMiner Studio* memiliki kurang lebih 500 operator *data mining*, termasuk operator *input*, *output*, *data preprocessing* dan visualisasi. *RapidMiner Studio* ini diprogram menggunakan bahasa *java* sehingga dapat bekerja pada semua sistem operasi.

METODE PENELITIAN

Penelitian ini menggunakan algoritma *k-means* untuk pengelompokan data alumni berdasarkan IPK, prodi, tahun lulus, angkatan dan jenis pekerjaan. Data yang diolah adalah data calon mahasiswa baru Universitas Muhammadiyah Yogyakarta (UMY).

Lokasi Penelitian

Lokasi penelitian dilaksanakan di Laboratorium Teknik Informatika Universitas Muhammadiyah Yogyakarta yang beralamat di Jalan Lingkar Selatan, Kasihan, Bantul, Yogyakarta.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Sumber data utama yang digunakan dalam penelitian ini adalah *database* alumni UMY dengan jumlah keseluruhan data ada 688 *record* dari angkatan 1998 sampai dengan tahun 2014. Tahun lulus 2003 sampai dengan tahun lulus 2016. Data yang diterima dalam bentuk tabel pada *microsoft excel* sehingga mempermudah untuk dilakukan pembersihan data atau memfilter data. Data ini di dapatkan melalui perijinan kepada pihak kampus UMY di bawah wewenang Biro Sistem Informasi (BSI).

Penyeleksian Data

Data alumni disaring terlebih dahulu dan diambil beberapa atribut dari tabel untuk dianalisis. Terdapat enam atribut yang digunakan yaitu, nama, prodi, IPK, tahun lulus, angkatan, dan jenis pekerjaan. Pengambilan data berdasarkan atribut yang digunakan seperti pada gambar 4.2 dan penyeleksian data menggunakan *sort & filter* yang ada pada *vitur excel*.

Algoritma K-Means

Dalam menggunakan algoritma *k-means* akan melakukan pengulangan tahapan hingga terjadi kestabilan. Peneliti melakukan tiga kali iterasi dalam melakukan pengujian menggunakan algoritma *k-means* pada data alumni dengan tahapan sebagai berikut.

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*. Kelompok *cluster* yang dibuat adalah 8 kelompok agar terlihat jarak antara titik *centroidnya*.
2. Penentuan nilai *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mangacu pada rumus *euclidean*.

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

3. Dilakukan pengelompokkan *centroid* sesuai dengan hasil dari jarak antar *centroid* tersebut. Hasilnya digunakan untuk penentuan kelompok *clustering*.

Pengujian Software RapidManer

1. Nilai *cluster centroid* seperti pada gambar 3.

2. Grafik *clustering* dari calon mahasiswa baru dalam memilih jurusan seperti pada gambar .

Tabel 1. Penentuan Frekuensi Dalam Menentukan Jarak *Centroid*.

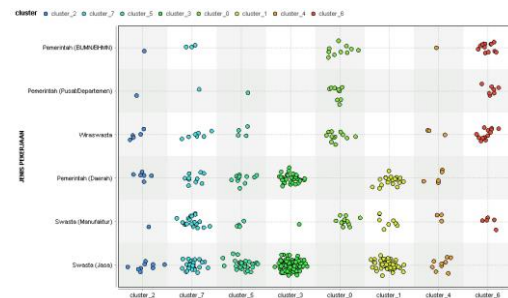
Centroid					
C0	16	3,70	2011	2007	1
C1	5	2,81	2012	2005	6
C2	10	3,00	2012	2006	1
C3	3	3,04	2011	2007	4
C4	4	2,98	2012	2007	3
C5	10	3,29	2012	2007	4
C6	9	3,50	2012	2007	1
C7	3	2,99	2011	2007	2

Tabel 2. Jarak antara *centroid*.

C0	C1	C2	C3	C4	C5	C6	C7
9,53	6,65	4,63	7,71	6,81	6,32	5,32	7,17
6,33	6,28	2,50	7,08	6,10	1,43	2,45	7,09
0,00	12,32	6,20	13,36	12,23	6,79	7,07	13,06
13,02	5,84	7,14	3,00	2,45	7,69	6,10	1,00
13,04	5,04	7,23	2,04	1,80	7,35	6,16	0,47
13,36	3,61	7,75	0,00	1,73	7,08	6,80	2,00

Atribus	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5	cluster_6	cluster_7
KODE PRODI	1.662	6.816	9	2.453	1.095	15	5.658	10.451
IPK	3.542	3.302	2.828	3.426	3.033	3.403	3.353	3.299
TAHUN LULUS	2012.877	2013.992	2012.304	2012.633	2012.967	2013.090	2013.474	2012.876
ANGKATAN	2008.795	2008.885	2004.609	2008.327	2004.905	2008.820	2009.053	2007.873
KODE JENIS PE.	4.341	1.425	2.391	1.247	2.143	1.720	4.579	2.211

Gambar 3. Hasil *Cluster Centroid* Dan *Cluster Instances* Dengan *Rapidminer Studio*.



Gambar 4. Hasil grafik *clustering*.

KESIMPULAN

Setelah diimplementasikan metode *clustering* menggunakan algoritma *k-means* pada data alumni, maka dapat disimpulkan bahwa :

1. Hasil analisis ini jenis pekerjaan yang paling dominan adalah swasta (jasa), dengan alumni paling banyak dari program studi Ilmu Hubungan Internasional, dengan rata-rata tahun lulus 2012, IPK yang sangat dominan antara 3,00 – 3,5.
2. Analisis ini dapat menjadi acuan universitas untuk melakukan strategi pemberdayaan mahasiswa terhadap dunia pekerjaan.

DAFTAR PUSTAKA

Angga Ginanjar Maburur, Lubis Riani. 2012. "Penerapan Data Mining Untuk

- Memprediksi Kriteria Nasabah Kredit.” *Jurnal Komputer dan Informatika (KOMPUTA)* 1 (1): 53–57.
- Anggraeni, Hapsari Dita, Ragil Saputra, dan Beta Noranita. 2013. “APLIKASI DATA MINING ANALISIS DATA TRANSAKSI PENJUALAN OBAT MENGGUNAKAN ALGORITMA APRIORI (Studi Kasus di Apotek Setya Sehat Semarang).” *Jurnal Masyarakat Informatika* 4 (7). <https://doi.org/10.14710/jmasif.4.7.1-8>.
- Asroni, Ronald Adrian, Asroni. 2015. “Penerapan Metode K-Means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang.” *Jurnal Ilmiah Semesta Teknika* 18 (1): 76–82. <https://doi.org/10.1038/hdy.2009.180>.
- Carvalho, M.J.; dan P. Melo-Goncalves. 2016. “Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation.” *Physics and Chemistry of the Earth, Parts A/B/C* 94 (Agustus): 22–28. <https://doi.org/10.1016/J.PCE.2016.05.001>.
- Fina Nasari, dan Surya Surya Darma. 2015. “Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru.” *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, 73–78.
- Ji, Jinchao, Wei Pang, Yanlin Zheng, Zhe Wang, dan Zhiqiang Ma. 2015. “A novel artificial bee colony based clustering algorithm for categorical data.” *PLoS ONE* 10 (5): 1–17. <https://doi.org/10.1371/journal.pone.0127125>.
- Josi Aranda, Josi, dan Wirda Wirda Astari Galvani Natasya. 2016. “PENERAPAN METODE K-MEANS CLUSTER ANALYSIS PADA SISTEM PENERAPAN METODE K-MEANS CLUSTER ANALYSIS PADA SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN KONSENTRASI UNTUK,” no. January 2018.
- Linda Maulida, Linda. 2018. “Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means” 2 (3): 167–74.
- Mudjihartono, Paulus, Thitipong Tanprasert, dan Rachsuda Jiamthapthaksin. 2016. “Clustering analysis on alumni data using abandoned and Reborn Particle Swarm Optimization.” *2016 8th International Conference on Knowledge and Smart Technology, KST 2016*, 22–26. <https://doi.org/10.1109/KST.2016.7440500>.
- Ong, Johan Oscar, Ong. 2013. “Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university.” *Jurnal Ilmiah teknik Industri* vol.12, no (juni): 10–20.
- Santosa; Samanda. 2017. “Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia.” *Jurnal Sains dan Seni Pomits* 6 (2): A51–56.
- Wardhani, Anindya Khrisna, W. 2016. “Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan.” *Jurnal Transformatika* 14 (1): 30–37.