



ANALISIS PENENTUAN PENERIMA BEASISWA POLITEKNIK POS INDONESIA MENGGUNAKAN CLUSTERING K-MEANS

Diki Wahyu Nugraha¹, Rd.Nuraini Siti Fathonah²

¹⁾ Bidang Kemahasiswaan, Alumni dan Kerjasama Politeknik Pos Indonesia, Jl. Sari Asih No. 54 Kode Pos 40151 Bandung, Jawa Barat, Indonesia

²⁾ Prodi/Jurusan D4 Teknik Informatika Politeknik Pos Indonesia, Jl. Sari Asih No. 54 Kode Pos 40151 Bandung, Jawa Barat, Indonesia

¹⁾ wahyunugrahadiki@gmail.com, ²⁾ nuraini@poltekpos.ac.id

INFO ARTIKEL

Diterima 15 Mei 2020

Direvisi 28 Juli 2020

Diterima 10 September 2020

Kata Kunci

Clustering K-Means, K-Means,
beasiswa, manajemen pendidikan,
administrasi pendidikan

Keywords

Clustering K-Means, K-Means,
scholarship, education management,
education administration

ABSTRAK

Proses Pelayanan Kemahasiswaan, Alumni, dan Kerjasama merupakan bagian tak terpisahkan dari administrasi POLITEKNIK POS INDONESIA (POLTEKPOS). Permasalahan yang terjadi saat ini adalah data-data yang ada di kemahasiswaan mengalami kesulitan dalam perhitungan sebuah data-data kemahasiswaan yang telah masuk ke dalam sistem. Maka dari itu dibuatlah suatu analisis dengan metode *K-means* dengan pengelompokkan data berdasarkan hasil pemilihan calon penerima beasiswa yang ada di kabag.kemahasiswaan untuk mahasiswa POLITEKNIK POS INDONESIA yang nantinya data akan diolah berdasarkan nilai *centroid* yang di dapat dalam masing-masing data yang telah di uji dalam metode *k-means clustering* data yang telah dianalisis pada bagian kemahasiswaan.

ABSTRACT

The process of Student Services, Alumni, and Cooperation is an inseparable part of the administration of POLITEKNIK POS INDONESIA (POLTEKPOS). The problem that occurs at this time is that the existing student data has difficulty in calculating student data that has been entered into the system. Therefore, an analysis was made using the *K-means* method by grouping data based on the results of the selection of prospective scholarship recipients in the head of student affairs for POLITEKNIK POS INDONESIA students which later the data will be processed based on the centroid value obtained in each of the data that has been collected. test in the *k-means clustering* method of data that has been analyzed in the student section.

* Corresponding author at:

Bidang Kemahasiswaan, Alumni dan Kerjasama
Politeknik Pos Indonesia, Jl. Sari Asih No. 54 Kode Pos 40151 Bandung, Jawa Barat, Indonesia.
E-mail address: wahyunugrahadiki@gmail.com

Published by School of Communication & Business, Telkom University.

1. Pendahuluan

Pelayanan Kemahasiswaan, Alumni, dan Kerjasama merupakan lembaga yang berada di bawah manajemen rektoran Politeknis Pos Indonesia (POLTEKPOS). Instansi ini mempunyai tugas untuk melakukan pelayanan akademik terhadap mahasiswa/mahasiswi yang ada di kampus POLTEKPOS, serta bertanggung jawab untuk memberikan informasi mengenai kegiatan kampus dan masalah akademik kepada mahasiswa/mahasiswi POLTEKPOS.

Dalam suatu lembaga pendidikan bidang kemahasiswaan, alumni dan kerjasama setiap tahun selalu membantu/menyalurkan informasi tentang adanya bantuan beasiswa kepada seluruh mahasiswa/mahasiswi POLTEKPOS. Beasiswa ini merupakan beasiswa yang disediakan untuk mahasiswa yang kurang mampu dan memiliki prestasi di bidang akademik maupun non akademik. Metode Fuzzy Logic Tsukamoto dari teknik clustering dapat membantu dalam mengklasifikasi mahasiswa yang berhak menerima beasiswa, mahasiswa yang dipertimbangkan menerima dan mahasiswa yang tidak berhak menerima beasiswa. Pada penelitian ini penulis yaitu pemanfaatan algoritma Fuzzy Logic Tsukamoto untuk menentukan penerima beasiswa dengan membagi data mahasiswa menjadi 2 cluster (Lengkong et al., 2015).

Namun permasalahan yang dihadapi oleh Bidang Kemahasiswaan, Alumni dan Kerjasama itu sendiri tidak adanya perhitungan dengan cepat untuk menentukan hasil penerima beasiswa karena pada dasarnya belum ditentukan perhitungan dengan cepat untuk menseleksi penerima calon beasiswa, dan terutama pihak beasiswa pun tidak bisa menanti lama menerima daftar data calon penerima beasiswa (Lengkong et al., 2015). Berdasarkan hasil penelitian, penulis menentukan sebuah pemilihan calon penerima beasiswa dengan menggunakan Fuzzy Logic Tsukamoto, sehingga dapat memudahkan penentu beasiswa untuk menentukan pemilihan calon mahasiswa/mahasiswi yang mendapatkan beasiswa secara waktu yang singkat. Adapun tujuan dari penelitian ini adalah membandingkan hasil cluster dari masing-masing format kriteria nilai dalam menentukan mahasiswa penerima beasiswa. Sehingga akan diketahui dari format kriteria nilai yang berbeda-beda, yang memiliki hasil cluster yang lebih baik.

Adapun identifikasi masalah yang melatarbelakangi penelitian ini adalah bagaimana cara menggunakan metode *Fuzzy Logic Tsukamoto* untuk memilih calon penerima beasiswa POLTEKPOS dengan waktu yang *real-time* dan cepat? Dan bagaimana cara menguji hasil dari *Fuzzy Logic Tsukamoto* untuk menentukan calon penerima beasiswa tersebut?

Adapun tujuan dari penelitian ini adalah sebagai berikut: untuk lebih memaksimalkan proses data secara berkelompok, untuk mengetahui berapa banyak data hasil calon penerima

beasiswa, untuk memberikan informasi calon penerima beasiswa sesuai dengan hasil yang telah di *clusteringkan* menggunakan metode *K-Means*. Manfaat yang diperoleh dari penelitian ini adalah menganalisis data beasiswa menggunakan metode *Fuzzy Logic Tsukamoto* dapat mempermudah mengelompokkan ketentuan penerima data calon beasiswa berdasarkan *Centroid*, selain itu dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan penyusunan data menggunakan metode *Fuzzy Logic Tsukamoto*. Ruang lingkup yang digunakan dalam penelitian ini adalah lingkup penelitian ini hanya dilakukan di lingkungan kampus bagian kemahasiswaan POLTEKPOS, dan data yang dijadikan input untuk analisis adalah data calon penerima beasiswa.

2. Tinjauan Pustaka

Algoritma *K-means* adalah algoritma pengelompokan iteratif sederhana, yang membagi sekumpulan data tertentu pada jumlah *cluster* (Defiyanti et al. 2017). Kelebihan Algoritma *clustering K-means* adalah lebih cepat mengelompokkan data serta mengumpulkan data yang besar secara cepat dan efisien. Pengelompokan *K-Means* bertujuan untuk membagi pengamatan ke dalam *cluster* k dimana setiap observasi menuju *cluster* dengan mean terdekat, berfungsi sebagai *prototype cluster*. Langkah-Langkah pada Algoritma *K-Means* adalah:

1. Menentukan jumlah *cluster* k
2. Menentukan *centroid* dari masing-masing *cluster*. Penentuan *centroid* dapat dilakukan dengan mengambil nilai secara random pada tiap *cluster*.
3. Menghitung jarak tiap data ke titik pusat *cluster/centroid* dengan menggunakan rumus jarak *Euclidean*, kemudian pindahkan titik data ke dalam kelompok dan kalkulasi ulang *centroid* baru. Hitung *centroid* ini berdasarkan rata-rata.
4. Ulangi langkah 3 hingga pusat cluster tidak berubah lagi.
5. Jika pusat *cluster* tidak berubah lagi, maka proses *clustering* selesai.

Rumus yang ada di Algoritma k-means yaitu pada rumus (1) (2):

$$r(x_i, x_j) = \sqrt{\sum_{k=1}^N (x_{ik} - x_{jk})^2}$$

Dimana $r(x_i, x_j)$ adalah jarak antara x_i dan x_j . x_i dan x_j adalah atribut dari data tertentu, dimana i, j dan k bervariasi dari 1 sampai N . N adalah jumlah atribut dari data yang diberikan.

Weka adalah paket *learning machine tools* praktis. "WEKA" merupakan singkatan dari "Waikato Environment for Knowledge Analysis", yang dibuat di *Universitas Waikato, New*

Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mampu menyelesaikan masalah-masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan *machine learning*. Perangkat lunak ini ditulis dalam hirarki class java dengan menggunakan metode berorientasi objek dan dapat berjalan hampir di semua platform. WEKA mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda-beda. Tersedia implementasi algoritma pembelajaran *state of the art* yang dapat diterapkan pada dataset data, *klasifikasi*, *regresi*, *clustering*, aturan *asosiasi* dan *visualisasi*. Penggunaan yang dapat melakukan *preprocess* data, memasukkannya dalam sebuah skema pembelajaran, dan menganalisis classifier yang dihasilkan dan performanya, semua itu tanpa menulis kode program sama sekali. Contohnya penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke data set dan menganalisis hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performanya untuk dipilih (Kartawinata & Wardhana, 2013).

Analisis *Cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *Cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama. *Cluster-cluster* yang terbentuk memiliki *homogenitas internal* yang tinggi dan *heterogenitas eksternal* yang tinggi (Pradana & Amir, 2016). Berbeda dengan teknik *multivariat* lainnya, analisis ini tidak mengestimasi *set variabel* secara empiris sebaliknya menggunakan *set variabel* yang ditentukan oleh peneliti itu sendiri.

Menurut Zhang et al. (2016), analisis *cluster* adalah membandingkan objek berdasarkan *set variabel*, hal inilah yang menyebabkan para ahli mendefinisikan *set variabel* sebagai tahap kritis dalam analisis *cluster*. *Set variabel cluster* adalah suatu *set variabel* yang merpresentasikan karakteristik yang dipakai objek-objek (Wardhana et al., 2016). Solusi *Analisis Cluster* bersifat tidak unik, anggota *cluster* untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi cluster secara keseluruhan bergantung pada *variabel-variabel* yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan *variabel-variabel* yang relevan dapat mempengaruhi substansi hasil analisis cluster. Pada tulisan ini penulis menggunakan metode *K-Means Cluster Analysis* sebagai solusi untuk pengklasifikasian karakteristik dari objek. Alasan penggunaan *algoritma K-Means* diantaranya ialah karena algoritma ini memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu *algoritma K-Means* ini tidak terpengaruh terhadap urutan objek. Permasalahan yang dikaji dalam

tulisan ini adalah bagaimana penggunaan *metode K-Means Cluster Analysis* dalam pengklasifikasian karakteristik suatu objek, tujuan yang ingin penulis capai mengenai *metode K-Means* adalah mengkaji *metode K-Means Cluster Analysis* dalam pengklasifikasian karakteristik berdasarkan *set variabel* yang dibentuk. Dengan menggunakan persamaan jarak penulis menghitung jarak objek terhadap centroid

Langkah-langkah analisisnya dimulai dari tentukan jumlah *cluster*, alokasikan objek ke dalam *cluster* secara *random*, hitung *centroid* sampel yang ada di masing-masing *cluster*, alokasikan masing-masing objek ke *centroid* terdekat, kembali ke langkah 3 apabila masih ada objek yang berpindah cluster atau masih ada perubahan nilai *centroid*, ada di atas nilai yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai yang ditentukan.

Sebagai bahan bacaan dalam tulisan ini penulis juga membandingkan beberapa karya ilmiah yang mengupas materi yang sama tentang *Analisis Cluster*, Rismawan dan Kusumadewi dalam penelitiannya mencoba membangun suatu sistem untuk mengelompokkan objek yang ada berdasarkan status gizi dan ukuran rangka dari objek yang diambil dengan memasukkan parameter kondisi fisik dari objek orang tersebut. Pengelompokan objek dilakukan dengan menggunakan *metode K-Means Cluster* yaitu mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas. Hasil dari penelitian ini terhadap 20 objek sampel diperoleh 3 kelompok mahasiswa berdasarkan nilai BMI (*Body Mass Index*) dan ukuran rangka, yaitu BMI normal dengan kerangka besar, BMI obesitas sedang dengan kerangka sedang, dan BMI obesitas berat dengan kerangka kecil.

K-Means Cluster Analysis merupakan salah satu *metode cluster analysis non-hirarki* yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain (Sugiharti & Muslim, 2016). Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang di set dalam proses *clustering*, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu *cluster* dan memaksimalkan variasi antar *cluster*. Metode *cluster* ini meliputi *sequential threshold*, *parallel threshold* dan *optimizing threshold* (Fakhri et al., 2016). *Sequential threshold* melakukan pengelompokan dengan terlebih dahulu memilih satu objek dasar yang akan dijadikan nilai awal *cluster*, kemudian semua *cluster* yang ada dalam jarak terdekat dengan *cluster* ini akan

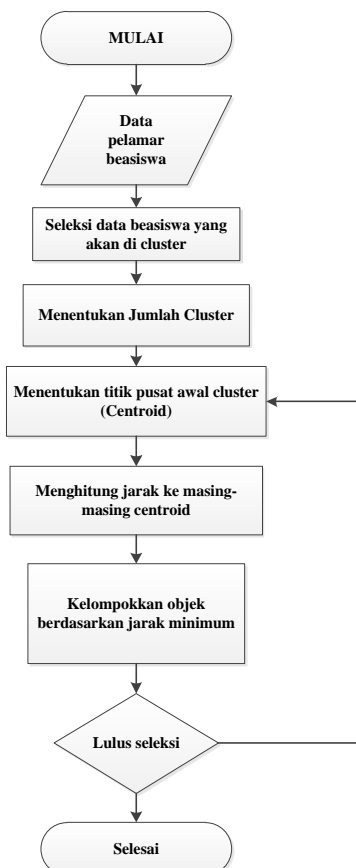
bergabung, lalu dipilih *cluster* kedua dan semua objek yang mempunyai kemiripan dengan *cluster* ini akan digabungkan, demikian seterusnya sehingga terbentuk beberapa *cluster* dengan keseluruhan objek yang terdapat didalamnya. Jika diberikan sekumpulan objek () maka algoritma *K-Means Cluster Analysis* akan mempartisi X dalam k buah *cluster*, setiap *cluster* memiliki centroid dari objek-objek dalam *cluster* tersebut. Pada tahap awal algoritma *K-Means Cluster Analysis* dipilih secara acak k buah objek sebagai *centroid*, kemudian jarak antara objek dengan *centroid* dihitung dengan menggunakan jarak *euclidian*, objek ditempatkan dalam *cluster* yang terdekat dihitung dari titik tengah *cluster*. *Centroid* baru ditetapkan jika semua objek sudah ditempatkan dalam *cluster* terdekat.

Proses penentuan *centroid* dan penempatan objek dalam *cluster* diulangi sampai nilai *centroid konvergen* (*centroid* dari semua *cluster* tidak berubah lagi). Secara umum metode *K-Means Cluster Analysis* menggunakan algoritma sebagai berikut, tentukan k sebagai jumlah *cluster* yang di bentuk. Untuk menentukan banyaknya *cluster k* dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*, tentukan k Bangkitkan k *Centroid* (titik pusat *cluster*) awal secara *random*. Penentuan *centroid* awal dilakukan secara *random* / acak dari objek-objek yang tersedia sebanyak *k cluster*, kemudian untuk menghitung *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut : $n \times v \quad n \quad \sum_{i=1}^n x_i = 1$; $i = 1, 2, 3, \dots$ dimana *centroid* pada *cluster* : objek ke-i n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*, hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid* penulis menggunakan *Euclidian Distance*. $(x_i - x_j)^2 + (y_i - y_j)^2$, $\sum_{i=1}^n = - - - - n \quad i \quad d \quad x$ $y \quad x \quad y \quad x_i \quad y_i$; $i = 1, 2, 3, \dots, n$ dimana x_i : objek x ke-i y_i : daya y ke-i n : banyaknya objek, alokasikan masing-masing objek ke dalam *centroid* yang paling terdekat. Untuk melakukan pengalokasian objek kedalam masing-masing *cluster* pada saat iterasi secara umum dapat dilakukan dengan dua cara yaitu dengan *hard k-means*, dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat *cluster* tersebut, cara lain dapat dilakukan dengan *fuzzy C-Means*, lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan . Ulangi langkah 3 jika posisi *centroid* baru tidak sama. Pengecekan konvergensi dilakukan dengan membandingkan *matriks group assignment* pada iterasi sebelumnya dengan *matrik group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma *k-means cluster analysis* sudah *konvergen*, tetapi jika berbeda maka belum *konvergen* sehingga perlu dilakukan iterasi berikutnya.

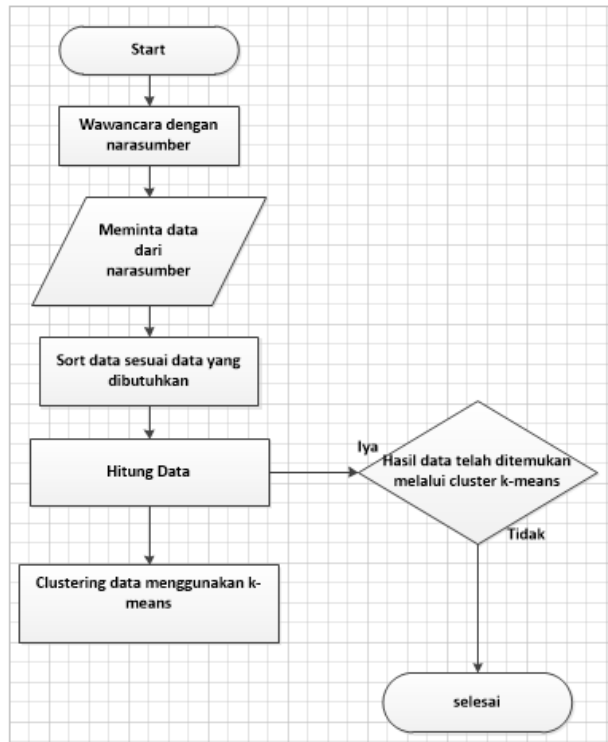
Pada penerapan *metode K-Means Cluster Analysis*, data yang bisa diolah dalam perhitungan adalah data numerik. Sedangkan data selain angka juga bisa diterapkan tetapi terlebih dahulu harus dilakukan pengkodean untuk mempermudah perhitungan suatu jarak/kesamaan karakteristik yg dimiliki dari setiap objek. Setiap objek dihitung kedekatan jaraknya berdasarkan karakter yang dimiliki dengan pusat *cluster* yang sudah ditentukan sebelumnya, jarak terkecil antara objek dengan masing-masing *cluster* merupakan anggota *cluster* yang terdekat. Setelah jumlah cluster ditentukan, selanjutnya dipilih sebanyak 3 objek secara acak sesuai jumlah *cluster* yang dibentuk sebagai pusat *cluster* awal untuk dihitung jarak kedekatannya terhadap semua objek yang ada.

3. Metode

Metode penelitian atau metode ilmiah adalah prosedur atau langkah-langkah dalam mendapatkan pengetahuan ilmiah atau ilmu (Pradana, 2015). Jadi, metode penelitian adalah cara sistematis untuk menyusun ilmu pengetahuan. Sedangkan teknik penelitian adalah cara untuk melaksanakan metode penelitian. Metode penelitian biasanya pada bentuk-bentuk penelitian. berikut langkah-langkah yang akan penulis jelaskan dalam tahapan-tahapan proses penelitian :



Gambar 1. Flowmap rancangan penentuan *cluster* penerima calon beasiswa menggunakan metode *k-means*.



Gambar 2. Flowmap rancangan wawancara untuk mengambil data kemudian dicluster untuk menentukan hasil data yang dipilih dalam calon penerima beasiswa.

4. Hasil dan Pembahasan

Dalam tahap ini akan dianalisis sumber-sumber permasalahan yang muncul di dalam bidang kemahasiswaan POLTEKPOS. Berdasarkan penelitian yang telah dilakukan terdapat permasalahan yaitu belum adanya perhitungan data calon penerima beasiswa pada setiap data pelayanan beasiswa dengan menentukan penilaian pada sistem menggunakan metode *K-Means*. Pengelompokan data beasiswa pada pelayanan kemahasiswaan menggunakan metode *K-Means* berguna untuk membantu perusahaan dalam mengetahui informasi dengan cepat dan akurat. Dengan mengelompokkan data beasiswa pelayanan kemahasiswaan, pihak staff dapat secara tepat memperkirakan atau melakukan penentuan keputusan sesuai dengan *cluster* yang telah ditentukan.

Data yang diperlukan dalam penelitian ini diperoleh melalui bagian kemahasiswaan POLTEKPOS yang berupa data pelayanan kemahasiswaan. Data-data tersebut berisi data beasiswa, penerimaan kerja sebelum wisuda namun dalam penelitian ini hanya menggunakan sampel data beasiswa dan penerimaan bekerja sebelum wisuda karena memiliki kriteria 2 nilai

dan beberapa bilangan data saja yang digunakan, seperti IPK, Pendapatan Orangtua, Jumlah Tanggungan. Berikut ini adalah data penerimaan beasiswa.

Tabel 1. Data Kandidat Mahasiswa Penerima Beasiswa

No	Kode Mahasiswa	Gaji Ortu	IPK	Jumlah Tanggungan
1	A	917000	2,33	5
2	B	1500000	3,41	4
3	C	1460000	1,9	4
4	D	1500000	2,92	4
5	E	2400000	3,52	5
6	F	2577850	3,15	4
7	G	3,37	3,37	5
8	H	2,90	2,90	6
9	I	3,23	3,23	3
10	J	3,01	3,01	6
11	K	2,43	2,43	7
12	L	2,85	2,85	10
13	M	3,05	3,05	11
14	N	2,68	2,68	11
15	O	3,10	3,10	12

Agar data di atas dapat diolah dengan menggunakan metode *k-means clustering*, maka data yang berjenis data nominal seperti IPK dan pendapatan orangtua, jumlah tanggungan harus diinisialisasikan terlebih dahulu dalam bentuk angka. Untuk melakukan transformasi data dilakukan dengan langkah-langkah sebagai berikut:

1. Pada data Pendapatan orangtua terlebih dulu dilakukan inisialisasi dalam bentuk angka.
2. Kemudian data IPK tersebut diurutkan dari yang terbesar berdasarkan nilai data penerimaan beasiswa yang berasal dari IPK tersebut.
3. Setelah itu Tanggungan Ortu yang memiliki nilai terbesar diberi inisial dengan angka 1 dan Tanggungan Ortu yang memiliki nilai terbesar kedua diberi inisial dengan angka 2, begitu seterusnya hingga *destination* dengan nilai paling sedikit.

Setelah semua data beasiswa dan data pekerjaan sebelum wisuda dalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-Means*

Clustering. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan mejadi dua *cluster*.
2. Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara *random* dan didapat titik pusat dari setiap *cluster* dapat dilihat pada table dibawah ini:

Tabel 2. Titik Pusat Awal

Titik Pusat Awal	A	B	C
C0	1,9	1460000	4
C1	3,37	4359900	5

Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *hard k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*.

Tabel 3. Iterasi-1

No	IPK	Gaji Ortu	Jumlah Tanggungan	Dc0	Dc1	C0	C1
1	2,33	917000	5	543000	3442900	2,33	0
2	3,41	1500000	4	40000	2859900	3,41	0
3	1,9	1460000	4	0	2899900	1,9	0
4	2,92	1500000	4	40000	2859900	2,92	0
5	3,52	2400000	5	940000	1959900	3,52	0
6	3,15	2577850	4	1117850	1782050	3,15	0
7	3,37	4359900	5	2899900	0	3,37	1
8	2,90	5407767	6	3947767	1047867	2,90	1
9	3,23	4601200	5	3141200	241300	3,23	1
10	3,01	4818400	5	3358400	458500	3,01	1
11	2,43	5147000	4	3687000	787100	2,43	1
12	2,85	3400000	6	1940000	959900	2,85	1
13	3,05	4900000	11	3440000	540100	3,05	1

14	2,68	3200000	12	1740000	1159900	2,68	1
15	3,10	5000000	13	3540000	640100	3,10	1

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.

Tabel 4. Titik Pusat Cluster Baru

Titik Pusat Baru	D0	D1	C0	C1
C0	3	1725808	4	3
C1	2,2911	4537140,7778	7,4444	2,2911

Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

Dalam penelitian ini, iterasi *clustering* data mahasiswa terjadi sebanyak 2 kali iterasi. Pada iterasi ke-2 ini, titik pusat dari setiap *cluster* sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

Tabel 5. Iterasi-2

No	IPK	Gaji Ortu	Jumlah Tanggungan	Dc0	Dc1	C0	C1
1	2,33	917000	5	543000	3442900	2,33	0
2	3,41	1500000	4	40000	2859900	3,41	0
3	1,9	1460000	4	0	2899900	1,9	0
4	2,92	1500000	4	40000	2859900	2,92	0
5	3,52	2400000	5	940000	1959900	3,52	0
6	3,15	2577850	4	1117850	1782050	3,15	0
7	3,37	4359900	5	2899900	0	3,37	1
8	2,90	5407767	6	3947767	1047867	2,90	1
9	3,23	4601200	5	3141200	241300	3,23	1
10	3,01	4818400	5	3358400	458500	3,01	1
11	2,43	5147000	4	3687000	787100	2,43	1
12	2,85	3400000	6	1940000	959900	2,85	1
13	3,05	4900000	11	3440000	540100	3,05	1
14	2,68	3200000	12	1740000	1159900	2,68	1
15	3,10	5000000	13	3540000	640100	3,10	1

Berdasarkan hasil pengelompokkan data menggunakan metode *k-means clustering*, di dapatkan hasil *clustering* hingga iterasi ke-2, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*. Hasil dari *clustering* tersebut seperti pada tabel di bawah ini :

Tabel 6. Hasil Cluster 1

NO	NAMA MAHASISWA	Jumlah IPK
1.	A	2,33
2.	B	3,41
3.	C	1,9
4.	D	2,92
Dengan rata-rata nilai IPK		2,9-3
Dengan rata-rata nilai Pendapatan Ortu		146 - 435
Dengan rata-rata nilai Jumlah Tanggungan		4 - 5

Dapat diketahui pada tabel 5.6 hasil *cluster* 1 calon penerima beasiswa terdiri dari 15 pelamar beasiswa BIDIK MISI A dengan IPK = 2,33, B dengan IPK = 3,41, C dengan IPK 1, , D dengan IPK 2,92. Dengan rata-rata banyaknya nilai Ipk 2,9 – 3, dengan rata-rata nilai pendapatan Ortu 1.460.000 – 4.350.000, dengan rata-rata nilai jumlah tanggungan 4-5 anak.

Tabel 7. Hasil Cluster 2

Cluster 2 data calon penerima beasiswa terdiri dari 4 data berasal dari data beasiswa BIDIK MISI :

No	Nama Calon Penerima Beasiswa	jumlah IPK
1.	E	3,01
2.	F	2,43
3.	G	2,85
4.	H	3,05
Dengan rata-rata nilai IPK		2,9-3
Dengan rata-rata nilai Pendapatan Ortu		146 - 435
Dengan rata-rata nilai Jumlah Tanggungan		4-5

Dapat diketahui pada tabel 5.6 hasil *cluster* 1 cabang HLP terdiri dari 15 pelamar beasiswa BIDIK MISI E dengan IPK = 23,01, Dika dengan IPK = 2,43, F dengan IPK 2,85, , G dengan IPK

3,05. Dengan rata-rata banyaknya nilai Ipk 2,43 – 3,05, dengan rata-rata nilai pendapatan Ortu 1.460.000 – 4.350.000, dengan rata-rata nilai jumlah tanggungan 4-5 anak.

Setelah itu, diadakan tahap pengkajian. Dalam pengkajian ini, peneliti akan melakukan kembali pengkajian pada data yang dilakukan oleh peneliti. Pengkajian yang dilakukan merupakan hasil pengumpulan data yang telah dilakukan untuk mendukung analisis dari penelitian ini. Sebelumnya, dijelaskan bahwa peneliti mengambil data dengan menggunakan wawancara kepada pihak internal atau *stakeholders* yang ada di POLTEKPOS terkait dengan proses atau tahapan dari metodologi penelitian yang digunakan yaitu Metode *K-MEANS*.

Data primer yang didapat dari penelitian ini merupakan data yang didapatkan dari hasil wawancara secara langsung dan pengambilan data calon pendaftar penerima beasiswa di POLTEKPOS.

Tabel 8. Data Primer Pendaftar Beasiswa

Nama	IPK	Gaji Orang Tua	Jumlah Tanggungan
A	2,33	917	5
B	3,41	150	4
C	1,9	146	4
D	2,92	150	4
E	3,52	240	5
F	3,15	250	4
G	3,37	435	5
H	2,90	540	6
I	3,23	460	5
J	3,01	480	5
K	2,43	550	4
L	2,85	340	6
M	3,05	490	3
N	2,68	320	2
O	3,10	500	3

Sebelumnya, sudah diperoleh hasil hasil pengelompokkan data menggunakan metode *k-means clustering*, didapatkan hasil *clustering* hingga iterasi ke-2, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*.

Tabel 9. Hasil akhir *clustering* iterasi-2

Titik pusat awal C1			Titik Pusat C2			DC0	DC1	Cluster
1,9	146	4	3,37	435	5	771	482	1
1,9	146	4	3,37	435	5	4	285	0

1,9	146	4	3,37	435	5	0	289	0
1,9	146	4	3,37	435	5	4	285	0
1,9	146	4	3,37	435	5	94	195	0
1,9	146	4	3,37	435	5	104	185	0
1,9	146	4	3,37	435	5	289	0	1
1,9	146	4	3,37	435	5	394	105	1
1,9	146	4	3,37	435	5	314	25	1
1,9	146	4	3,37	435	5	334	45	1
1,9	146	4	3,37	435	5	404	115	1
1,9	146	4	3,37	435	5	194	95	1
1,9	146	4	3,37	435	5	344	55	1
1,9	146	4	3,37	435	5	174	115	1
1,9	146	4	3,37	435	5	354	65	1

Kemudian pada tabel 10 terlihat yang ditarik dari data di tabel 9 diuraikan menjadi sebuah informasi yang akurat.

Tabel 10. Hasil Cluster 1 dan Cluster 2

Hasil Cluster 1		
No	NAMA Calon Penerima Beasiswa	jumlah IPK
1.	A	2,33
2.	B	3,41
3.	C	1,9
4.	D	2,92
5.	E	3,52
6.	F	3,15
Dengan rata-rata nilai IPK		2,9-3
Dengan rata-rata nilai Pendapatan Ortu		146 - 435
Dengan rata-rata nilai Jumlah Tanggungan		4 - 5

Hasil Cluster 2		
Nomor	Nama Calon Penerima Beasiswa	jumlah IPK
7	G	3,37
8	H	2,90
9	I	3,23
10	J	3,01
11	K	2,43
12	L	2,85
13	M	3,05

14	N	2,68
15	O	3,10
Dengan rata-rata nilai IPK		2,9-3
Dengan rata-rata nilai Pendapatan Ortu		146 - 435
Dengan rata-rata nilai Jumlah Tanggungan		4 - 5

Dapat diketahui hasil *cluster* 1 calon penerima beasiswa terdiri dari 6 mahasiswa diantaranya : terdiri dari 6 nama beasiswa BIDIK MISI A dengan IPK = 2,33,B dengan IPK = 3,41,C dengan IPK 1, ,D dengan IPK 2,92.Dengan rata-rata banyaknya nilai Ipk 2,9 – 3,dengan rata-rata nilai pendapatan Ortu 1.460.000 – 4.350.000,dengan rata-rata nilai jumlah tanggungan 4-5 anak.

Metode K-means Clustering adalah metode pengelompokan yang bertujuan mengelompokan individu sedemikian hingga jarak setiap individu ke pusat kelompok dalam satu kelompok adalah minimum (Hatono et al., 2018). Proses *clustering* dengan menggunakan algoritma *K-Means* memiliki langkah-langkah sebagai berikut :

1. Tentukan jumlah *cluster* (sebanyak *K cluster*)
2. Menentukan *centroid* setiap *cluster* secara acak.
3. Menghitung jarak setiap objek ke *centroid*
4. Memasukkan objek ke *centroid* terdekat.
5. Perbaharui kembali *centroid* setiap *cluster*.
6. Kembali ke langkah 3, apabila masih ada data yang berpindah *cluster*.

Jarak antara data dihitung dengan menggunakan rumus K-Means sebagai berikut :

$$=(((B3-E3)^2+(C3-F3)^2+(D3-G3)^2)^{0,5})$$

Dimana :

x = Objek Data.

y = Data *Cluster*.

m= Jumlah Atribut.

Untuk implementasinya, kita bisa bayangkan kasus sederhananya seperti ini; andaikan suatu jurusan di suatu kampus pada gelombang pertama memiliki lulusan sebanyak 15 mahasiswa, dan kita ingin membagi ke-15 mahasiswa tersebut menjadi dua kelompok sehingga anggota satu kelompok memiliki karakteristik yang jauh berbeda dengan kelompok yang lain

(untuk kasus yang ini karakteristik yang dimaksud adalah berdasarkan IPK dan Gaji Orang tua, Tanggungan Orang tua).

Berikut adalah data mahasiswa calon penerima beasiswa :

Tabel 11. Data mahasiswa calon penerima beasiswa

Nama	IPK	Gaji Orang Tua	Jumlah Tanggungan
A	2,33	917	5
B	3,41	150	4
C	1,9	146	4
D	2,92	150	4
E	3,52	240	5
F	3,15	250	4
G	3,37	435	5
H	2,90	540	6
I	3,23	460	5
J	3,01	480	5
K	2,43	550	4
L	2,85	340	6
M	3,05	490	3
N	2,68	320	2
O	3,10	500	3

Setelah ditentukan clusternya maka cluster tersebut dihitung, ada berapa cluster di C0 dan ada berapa cluster di C1.

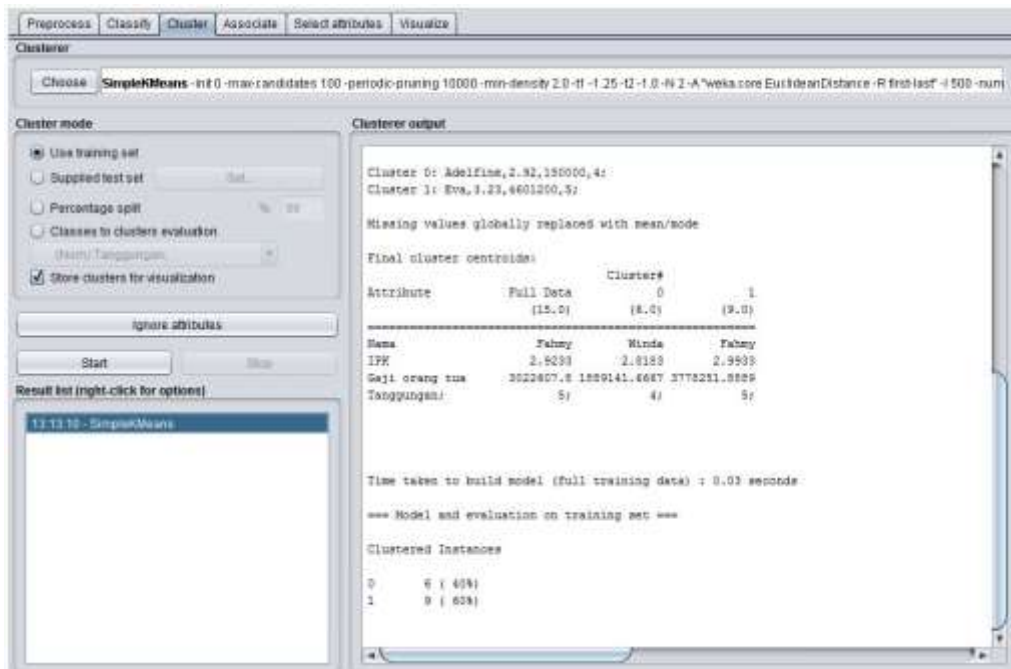
Tabel 12. Hasil cluster yang telah dihitung menggunakan rumus k-means

	A	B	C
C0	3	187	4
C1	2,2950	503,2000	4,4000

Hasil perhitungan pembagian cluster dengan rumus $=\text{SUM}(\text{jumlahkan cluster} + \text{kemudian bagi sesuai ketentuan cluster yang telah diperhitungkan})$. Rumus perhitungan cluster agar menetapkan hasil adalah sebagai berikut :

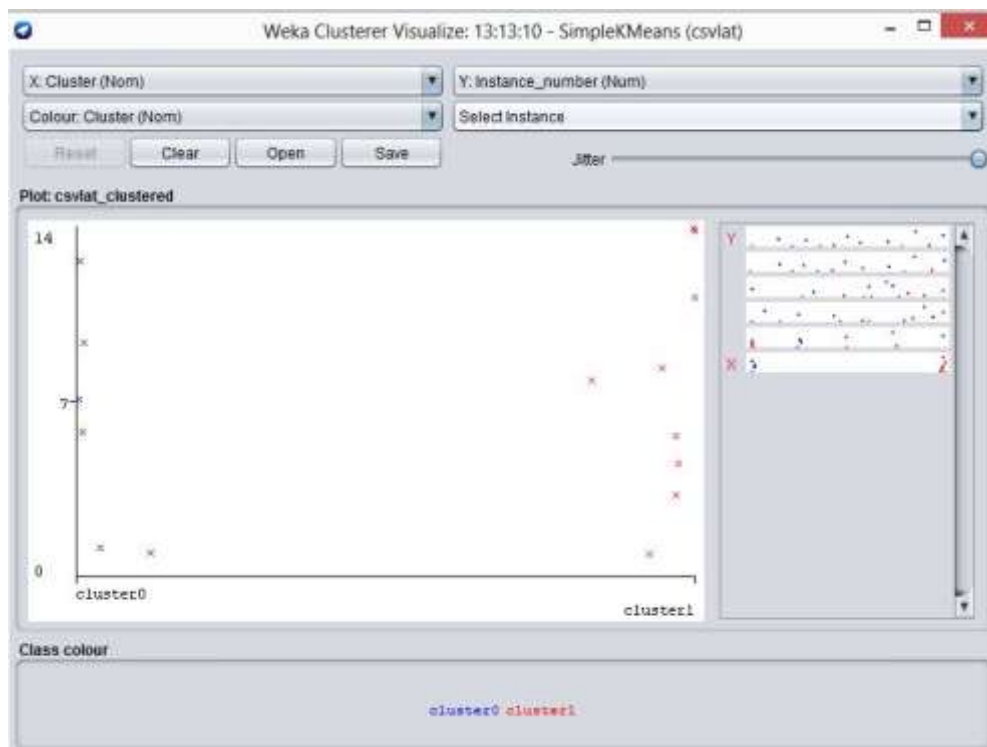
$$=\text{SUM}(B3-B7)/7$$

$$=\text{SUM}(B9-B15)/8$$



Gambar 3. Hasil *cluster centroids* dan *clustered Instances* dengan WEKA

Dapat dilihat pada Figur 3 adalah hasil *cluster centroids* dan *clustered Instances* menggunakan aplikasi WEKA terdiri dari 4 atribut yang digunakan yaitu Nama, IPK, Tanggungan dan gaji orangtua. iterasi sebanyak 2 kali, menampilkan hasil akhir rata-rata *full data*, rata-rata data berdasarkan *cluster 1*, rata-rata data berdasarkan *cluster 2* dan jumlah *cluster 0* = 6 dan jumlah *cluster 1* = 9.



Gambar 4. Hasil Grafik clustering dengan WEKA

Dapat dilihat figure 4 merupakan suatu grafik hasil *clustering* berdasarkan *Cluster 1* dan *Cluster 2*.

No.	1: Instance_number	2: Nama	3: IPK	4: Gaji orang tua	5: Tanggungan	6: Cluster
	Numeric	Nominal	Numeric	Numeric	Nominal	Nominal
1	0.0	Fahmy	2.33	917000.0	5;	cluster1
2	1.0	Winda	3.41	1500000.0	4;	cluster0
3	2.0	Vera	1.9	1460000.0	4;	cluster0
4	3.0	Adelfi...	2.92	150000.0	4;	cluster0
5	4.0	Suci	3.52	2400000.0	5;	cluster1
6	5.0	Dina	3.15	2577850.0	4;	cluster0
7	6.0	Annisa	3.37	4359900.0	5;	cluster1
8	7.0	Dewa...	2.9	5407767.0	6;	cluster1
9	8.0	Eva	3.23	4601200.0	5;	cluster1
10	9.0	Wisnu	3.01	4818400.0	5;	cluster1
11	10.0	Dika	2.43	5147000.0	4;	cluster0
12	11.0	Halima	2.85	3400000.0	6;	cluster1
13	12.0	Inka	3.05	4900000.0	12;	cluster1
14	13.0	Pipit	2.68	3200000.0	12;	cluster1
15	14.0	Much...	3.1	500000.0	13;	cluster0

Gambar 5. Hasil clustering dengan WEKA

Pada figure 5 dapat diketahui hasil *clustering* menggunakan aplikasi WEKA yang merupakan suatu objek sesuai dengan *cluster-cluster* yang ditentukan. Dengan demikian hasil pengujian menggunakan aplikasi WEKA, dapat diketahui persamaan - persamaan dengan analisis yang dilakukan dan dapat disimpulkan bahwa suatu analisis menggunakan *K-Means* tersebut setiap tahap-tahap atau proses yang dilakukan maupun perhitungannya adalah benar.

5. Kesimpulan

Kesimpulan yang dapat diambil dari hasil penentuan calon penerima beasiswa adalah dengan adanya analisis dan perancangan hasil penentuan calon penerima beasiswa menggunakan metode *clustering k-means* membantu proses perhitungan dan penentuan secara cepat dan tepat. Penelitian ini menghasilkan sebuah perhitungan calon penerima beasiswa menggunakan metode

clustering k-means yang dapat membantu dalam melaksanakan ketentuan calon penerima beasiswa. Selain itu juga membantu dalam melakukan proses perhitungan secara cepat menggunakan metode *k-means*.

Adapun saran-saran yang ingin disampaikan adalah perancangan ini dapat ditambah atau dikembangkan lagi sesuai dengan kebutuhan yang diinginkan.

DAFTAR PUSTAKA

- Defiyanti, S., Jajuli, M., & Rohmawati, N. (2017). Optimalisasi K-Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion. *Jurnal Nasional Teknologi dan Sistem Informasi*, 3(1), 211-218.
- Fakhri, M., Aditya, M., & Pradana, M. (2014). Factor analysis of work motivation using Maslow's hierarchy of needs: Case study on civil servants at Banten Province's Office of Agriculture and Livestock. In *International Conference on Emerging Trends in Academic Research* (pp. 258-274).
- Hartono, O S Sitompul, Tulus & E B Nababan. (2018). Optimization Model of K-Means Clustering Using Artificial Neural Networks to Handle Class Imbalance Problem. *The 2nd Annual Applied Science and Engineering Conference (AASEC 2017)*.
- Kartawinata, B. R., & Wardhana, A. (2013). Marketing strategies and their impact on marketing performance of Indonesian ship classification society. *International Journal of Science and Research (IJSR)*, 4(2), 69-74.
- Lengkong, S. P., Permanasari, A. E., & Fauziati, S. (2015). Implementasi Metode VIKOR Algoritma K-Means untuk Seleksi Penerima Beasiswa. *Proceedings of The 7 Th National Conference on Information Technology and Electrical Engineering*, 33, 107-112.
- Pradana, M. (2015). Pengaruh Gaya Kepemimpinan Terhadap Motivasi Karyawan di Ganesha Operation, Bandung. *Jurnal Studi Manajemen dan Bisnis*, 2(1), 24-39.
- Pradana, M., & Amir, N. (2016). Measuring E-Learning Effectiveness at Indonesian Private University. *International Journal of Environmental and Science Education*, 11(18), 11541-11556.
- Sugiharti, E., & Muslim, M.A. 2016. On-line Clustering of Lecturers Performance of Computer Science Department of Semarang State University Using K-Means Algorithm. *Journal of Theoretical and Applied Information Technology*, 83(1).

- Wardhana, A., Syahputra, S., & Kartawinata, B. R. (2017). Determinant Factors of Consumer Preferences in Indonesia Airlines Industry. *Jurnal Bisnis dan Manajemen*, 18(1), 11-20.
- Zhang, X., Jiang, J., Ge, B., & Yang, K. (2016). Group decision making for weapon systems selection with VIKOR based on consistency analysis. *Systems Conference (SysCon) 2016 Annual IEEE*, 1-6.