

# ANALISIS DAN IMPLEMENTASI FRAMEWORK CRISP-DM (*CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING*) UNTUK *CLUSTERING* PERGURUAN TINGGI SWASTA

Sri Erina Damayanti<sup>1</sup>, Sri Kuswayati<sup>2</sup>  
Prodi Teknik Informatika<sup>1,2</sup>

Sekolah Tinggi Teknologi Bandung, Jalan Soekarno Hatta No. 378 Bandung

[Erina@sttbandung.ac.id](mailto:Erina@sttbandung.ac.id)<sup>1</sup>

## Abstrak

Penelitian ini bertujuan untuk mengelompokkan Perguruan Tinggi Swasta (PTS), dengan harapan agar lebih tepat dalam melakukan strategi, pengawasan, dan kualitas perguruan tinggi harus ditingkatkan di Kopertis IV yang berada di wilayah Jawa Barat. *Clustering* adalah salah satu teknik data mining harus dapat memenuhi suatu standar proses yang telah disepakati. CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah standar proses data mining yang akan digunakan pada penelitian ini. CRISP-DM terbagi enam fase yang berkaitan satu sama lain. Adapun teknik yang digunakan yaitu *Clustering*, dan *software* yang digunakan SPSS 21.

Model yang dihasilkan dari penelitian ini dapat dimanfaatkan sebagai acuan pengelompokan Perguruan Tinggi Swasta (PTS), serta dapat memberikan informasi bagi masyarakat dan dapat digunakan sebagai dasar kebijakan pengembangan Perguruan Tinggi Swasta (PTS) dengan baik di Kopertis IV Wilayah Jawa Barat.

Kata kunci : *Clustering*, CRISP-DM, Perguruan Tinggi Swasta, SPSS

## Abstract

This study aims to classify Private Higher Education (PTS), with the hope that it is more appropriate to carry out the strategy, supervision, and quality of higher education institutions to be improved at Kopertis IV in the West Java region. Clustering is one of the data mining techniques that must be able to meet an agreed standard process. CRISP-DM (Cross Industry Standard Process for Data Mining) is a standard data mining process that will be used in this study. CRISP-DM is divided into six phases that are related to each other. The techniques used are Clustering, and the software used by SPSS 21.

The model produced from this study can be used as a reference for grouping Private Universities (PTS), and can provide information for the community and can be used as a basis for the development of a well- established private university in Kopertis IV West Java Region.

Keyword : Clustering, CRISP-DM, Private Universities, SPSS

## I. PENDAHULUAN

Bertambah banyaknya jumlah Perguruan Tinggi, khususnya Perguruan Tinggi Swasta (PTS), sangat berdampak pada persaingan yang sangat ketat, oleh karena itu setiap Perguruan Tinggi Swasta (PTS) harus meningkatkan kualitas pendidikannya dari berbagai faktor, misalnya faktor kualitas Manajemen, kualitas Sumber Daya Manusia, Akreditasi Institusi, dan lain sebagainya.

Menurut (Romi Wahyono, 2007), tercantumnya nama sebuah Perguruan Tinggi dalam pemeringkatan, terutama pada *top level*, menurut kalangan tertentu merupakan hal penting dan patut dilakukan. Pertimbangannya, antara lain masuknya suatu perguruan tinggi dalam top perguruan tinggi, merupakan langkah awal yang sangat baik agar dunia pendidikan tinggi mengakui keberadaannya, dapat mendorong citra perguruan tinggi, juga akan meningkatkan promosi, terutama dalam perolehan mahasiswa dan dukungan pihak luar dalam penelitian. Hal itu juga merupakan salah satu wujud nyata upaya menyukseskan program Pendidikan Nasional.

Kualitas Perguruan Tinggi dapat diketahui melalui mengelompokkan Perguruan Tinggi yaitu dengan menggunakan disiplin ilmu *data mining*. "*Data mining* merupakan analisis dari peninjauan kumpulan data untuk

menentukan hubungan yang tidak diduga dan meringkas data cara berbeda sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data" (Larose, 2006).

Teknik *data mining* yang digunakan untuk mengelompokkan Perguruan Tinggi Swasta (PTS) adalah menggunakan teknik *clustering*. Teknik *clustering* digunakan pada data mining untuk mengelompokkan objek-objek yang memiliki kemiripan dalam kelas atau segmen yang sama (Jain, 1999).

Berdasarkan uraian diatas, maka penulis akan melakukan penelitian yaitu: "Analisis dan Implementasi Framework Crisp-DM (*Cross Industry Standard Process For Data Mining*) Untuk *Clustering* Perguruan Tinggi Swasta", diharapkan dengan dilakukan penelitian ini dapat membantu memberikan informasi mengenai pengelompokkan Perguruan Tinggi Swasta (PTS) di Kopertis Wilayah IV Jawa Barat.

## II. TINJAUAN PUSTAKA

### A. Data Mining

*Data mining* merupakan proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, *data warehouse*, atau tempat penyimpanan data lainnya (Tan dkk, 2006). "*Data mining* sebagai disiplin ilmu yang

tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Kegiatan inilah yang menjadi garapan atau perhatian utama dari disiplin ilmu *data mining*” (Susanto, 2010).

Menurut (Sumanthi dan Sivandham, 2009), *data mining* juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam *database* yang sering disebut dengan istilah *Knowledge Discovery in Database* (KDD). KDD merupakan suatu area yang mengintegrasikan berbagai metode, yang meliputi statistik, basis data, kecerdasan buatan (*Artificial Intelligence*), *machine learning*, pengenalan pola (*Pattern Recognition*), pemodelan yang menangani ketidakpastian, visualisasi data, optimasi, Sistem Informasi Manajemen (SIM), dan sistem berbasis pengetahuan (*knowledge based-system*). Sebagai bagian dari proses yang ada di dalam KDD, maka *data mining* di dahului dengan proses pemilihan data, pembersihan data, *pre-processing*, dan transformasi data (Sumanthi dan Sivandham, 2009).

Sebagai bagian dari proses yang ada di dalam KDD, maka *data mining* di dahului dengan proses pemilihan data, pembersihan data, *pre-processing*, dan transformasi data (Sumanthi dan Sivandham, 2009). Ada tiga tahap penting dalam KDD (Tan dkk, 2006), adalah sebagai berikut:

#### 1. *Data preprocessing*

Proses ini bertujuan untuk mentransformasikan data *input* ke dalam format yang sesuai untuk kemudian dianalisa. Dalam tahap ini dilakukan proses penggabungan data dari berbagai sumber, pembersihan data untuk menghilangkan *noise data* dan data ganda, serta memilih atribut data yang diperlukan bagi proses *data mining*.

#### 2. *Data mining*

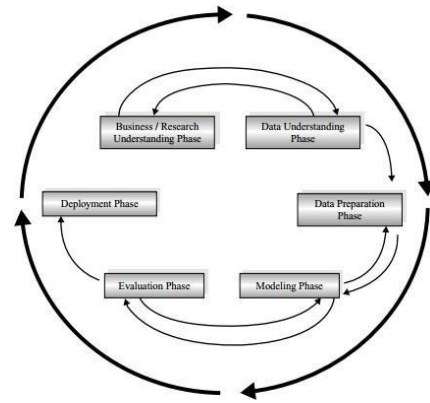
Proses ini bertujuan untuk mendapatkan pola-pola dan informasi yang tersembunyi di dalam basis data. Ada beberapa teknik yang dapat digunakan dalam *data mining* untuk mendapatkan pola-pola dan informasi tersembunyi, yaitu *classification*, *neural network*, *decision tree*, *genetic algorithm*, *clustering*, *OLAP* (*Online Analytical Processing*), dan *association rules*.

#### 3. *Postprocessing*

Proses ini bertujuan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan. Contoh dari proses ini adalah proses visualisasi, yaitu proses untuk menganalisa dan mengeksplorasi data dan hasil dari proses *data mining* dari berbagai sudut pandang.

### B. *CRISP-DM* (*Cross Industry Standard Process For Data Mining*)

*CRISP-DM* (*Cross Industry Standard Process for Data Mining*) yang dikembangkan tahun 1996 oleh analisis dari beberapa industri seperti standarisasi *Daimler Chrysler* (*Daimler-Benz*), SPSS, NCR. *CRISP-DM* menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian (Larose, 2006). Proses *CRISP-DM* terlihat pada gambar 1.



Gambar 1. Proses *Data Mining* *CRISP-DM* (Larose, 2006)

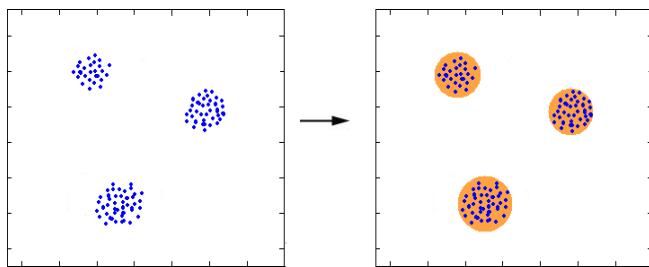
Proses *data mining* berdasarkan *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) terdiri dari enam fase (Larose, 2006), adalah sebagai berikut:

1. *Business Understanding Phase* (Fase Pemahaman Bisnis)
  - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
  - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
  - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. *Data Understanding Phase* (Fase Pemahaman Data)
  - a. Mengumpulkan data.
  - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
  - c. Mengevaluasi kualitas data.
  - d. Jika diinginkan, pilih sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan.
3. *Data Preparation Phase* (Fase Persiapan Data)
  - a. Siapkan dari data awal, kumpulkan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
  - b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
  - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
  - d. Siapkan data awal sehingga siap untuk perangkat permodelan.
4. *Modelling Phase* (Fase Pemodelan)
  - a. Pilih dan aplikasikan teknik pemodelan yang sesuai
  - b. Kalibrasi aturan model untuk mengoptimalkan hasil.
  - c. Perlu diperhatikan bahwa beberapa teknik mungkin digunakan untuk permasalahan *data mining* yang sama.
  - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
5. *Evaluation Phase* (Fase Evaluasi)
  - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
  - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.

- c. Menetapkan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
  - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. *Deployment Phase* (Fase penyebaran)
- a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
  - b. Contoh sederhana penyebaran: Pembuatan laporan.

**C. Clustering**

Pengelompokan (*clustering*) merupakan teknik yang sudah cukup dikenal dan banyak digunakan untuk mengelompokkan data/objek ke dalam kelompok data (*cluster*) sehingga setiap *cluster* memiliki data yang mirip dan berbeda dengan data yang berada dalam *cluster* lain. Jika diberikan himpunan data yang berjumlah terhingga, yaitu  $X$ , maka permasalahan *clustering* dalam  $X$  adalah mencari beberapa pusat *cluster* yang dapat memberikan ciri kepada masing-masing *cluster* dalam  $X$ . Gambar 2. menunjukkan contoh sederhana dari proses *clustering*, dimana 3 *cluster* di dalam data dapat teridentifikasi dengan mudah. Kriteria kemiripan yang digunakan dalam kasus ini adalah jarak (dalam kasus ini jarak geometris). Proses ini disebut *distance-based clustering*. Cara lain untuk melakukan *clustering* adalah *conceptual clustering*. Dalam *conceptual clustering*, objek dikelompokkan berdasarkan kecocokannya menurut konsep deskriptif.



Gambar 2. Proses Clustering (Jansen, 2007)

*Clustering* adalah studi formal metode dan algoritma untuk partisi atau pengelompokan. Analisis *clustering* tidak menggunakan pelabelan kategori sebelumnya. *Clustering* bersifat *unsupervised learning* atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi (Jain, 2009).

*Clustering* adalah alat penemuan mengungkapkan hubungan dan struktur di dalam data yang sebelumnya tidak jelas menjadi pengetahuan yang bermanfaat ketika ditemukan. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data/obyek ke dalam *cluster* (*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* adalah metode yang berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster* dan membuat jarak antar *cluster* sejauh mungkin. Ini berarti obyek dalam satu *cluster* sangat mirip satu sama lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain.

**III. ANALISIS DAN PERANCANGAN**

**A. Tahapan CRISP-DM**

Tahapan CRISP-DM terdiri dari enam fase yang harus

dilakukan, sebagaimana yang dijelaskan pada CRISP-DM (*Cross Industry Standard Process for Data Mining*).

1. Fase Pemahaman Bisnis

Pemahaman bisnis adalah tahap pertama dalam proses CRISP-DM, yaitu menentukan tujuan bisnis

2. Fase Pemahaman Data

Fase pemahaman data pada penelitian ini yaitu dimulai dari pengumpulan data, tahap pemahaman data ini akan dilakukan pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan bisnis. Apabila terdapat definisi variable yang tidak sesuai dengan tujuan penelitian, maka dilakukan perumusan masalah ulang, namun apabila definisi dari variable variable sudah sesuai dengan tujuan penelitian maka akan ditentukan variabel-variabel yang akan digunakan.

3. Fase Persiapan Data

Fase persiapan data awal merupakan tahap awal untuk melakukan perhitungan statistic deskriptif terhadap keseluruhan variabel yang ada.

4. Pemodelan

Fase pemodelan pada penelitian ini secara langsung melibatkan teknik data mining. Adapun teknik data mining yang digunakan adalah *clustering* dengan menggunakan algoritma *K-Means*.

Sebelum melakukan pengolahan data menggunakan teknik *clustering* dengan algoritma *K-Means*, maka diperlukan standarisasi data perlu dilakukan karena terdapat banyak jenis satuan pada data. Misalnya satuan kualitas SDM, jumlah prodi, rasio dosen tetap, jumlah mahasiswa, dan lain sebagainya. Perbedaan satuan yang mencolok dapat menyebabkan terjadinya bias pada analisis cluster sehingga data asli harus distandarasi sebelum bisa dianalisis. *Software* yang digunakan dalam proses standarisasi adalah *software SPSS 2.1*.

Jumlah *cluster* ( $K$ ) yang digunakan pada penelitian ini adalah lima kelompok (*cluster*), diharapkan agar mendapatkan hasil pengelompokan yang optimal dan hasil tersebut akan mewakili kategori kelompok yang terbentuk. Adapun hasil dari pengelompokan sekaligus jawaban dari rumusan masalah.

Proses *clustering* untuk penerapan algoritma *K-Means* pada pengelompokan 352 Perguruan Tinggi Swasta (PTS) yang berada di wilayah Kopertis IV, dimulai dari *initial cluster* dapat dilihat pada tabel I.

Tabel I. Initial Cluster Center

	Cluster				
	1	2	3	4	5
Zscore:	1.2807	-	1.2161	2.7130	2.3775
KUALITAS SDM	1	1.5840	9	8	7
Zscore:	1.3986	-	1.8330	2.7018	2.8104
KUALITAS MANAJEMEN	1	.77351	3	8	9
Zscore:	3.2947	-	6.9417	-	5.3787
KUALITAS PENELITIAN	6	.35227	8	.35227	7

N & PUBLIKASI					
Zscore:	1.4522	-	1.4522	1.4522	1.4522
AKREDITASI INSTITUSI	5	.68663	5	5	5
Zscore:	6.4812	-	3.6811	-	3.8458
JUMLAH PRODI	4	.10716	7	.43657	8

Untuk mendeteksi berapa kali iterasi yang dilakukan dalam proses *clustering* dari 352 objek yang diteliti, dapat dilihat pada tabel II.

Tabel II. Iteration History<sup>2</sup>

	Change in Cluster Centers				
	1	2	3	4	5
1	.000	.000	.000	4.189	5.874
2	.000	.000	.000	.056	1.214
3	.000	.000	.000	.031	.717
4	.000	.000	.000	.016	.359
5	.000	.000	.000	.000	.000

*Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 5. The minimum distance between initial center is 14.298.*

Pada tabel diatas ditampilkan proses iterasi yang terjadi. Pada proses iterasi tersebut dilakukan percobaan untuk mengubah *cluster* yang ada (initial). Tujuan iterasi tersebut adalah berusaha mencari pengelompokan yang paling tepat terhadap objek amatan. Pada tabel tersebut ditunjukkan bahwa dibutuhkan 5 iterasi untuk menempatkan objek pada *cluster* yang tepat. Jarak minimum antar pusat *cluster* yang terjadi dari hasil iterasi adalah 14.298. Setelah dilakukan 5 iterasi *cluster*, maka hasil akhir dari proses *clustering* dapat dilihat pada tabel III.

Tabel III. Final Cluster Centers

	Cluster				
	1	2	3	4	5
Zscore:	1.2807	-	1.2161	-	1.3774
KUALITAS SDM	1	1.5840	9	.0602	9
Zscore:	1.3986	-	1.8330	-	1.4994
KUALITAS MANAJEMEN	1	.77351	3	.0700	6
Zscore:	3.2947	-	6.9417	-	3.2203
KUALITAS PENELITIAN & PUBLIKASI	6	.35227	8	.1640	3
Zscore:	1.4522	-	1.4522	-	.99392
AKREDITASI INSTITUSI	5	.68663	5	.0481	6
Zscore:	6.4812	-	3.6811	-	3.2341
JUMLAH PRODI	4	.10716	7	.1651	0
				7	

Tabel diatas adalah hasil final dari 5 iterasi kelompok (*cluster*), data yang berada pada tabel diatas dapat dianalisis dan ditinterpretasikan. Analisis *cluster* pada data diatas dapat

dilakukan dengan dua metode. Metode pertama adalah dengan menjabarkan secara rinci tafsiran nilai Z dari masing masing *cluster*. Cara kedua dapat dilakukan dengan melakukan penafsiran secara umum, dimana sudah cukup memadai jika dikatakan “lebih dari rata rata (+)” atau “kurang dari rata rata (-)”. Berikut adalah analisis terhadap 5 kelompok (*cluster*) yang terbentuk pada data tabel di atas dapat didefinisikan sebagai berikut:

**a. Kelompok-1 (Cluster – 1)**

Dalam kelompok ini berisikan perguruan tinggi yang mempunyai kualitas sdm, kualitas manajemen, kualitas penelitian dan publikasi, akreditasi institusi, jumlah prodi yang lebih dari rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai positif yang terdapat pada table *final cluster centers* dalam keseluruhan *variable*. Dengan demikian dapat disimpulkan bahwa kelompok-1 ini merupakan pengelompokan dari perguruan tinggi swasta terbaik.

**b. Kelompok-2 (Cluster – 2)**

Karakteristik perguruan tinggi yang mengelompok pada kelompok-2 adalah hampir keseluruhan instrument penilai berada pada posisi dibawah rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai negatif yang terdapat pada table *final cluster centers*, namun jumlah mahasiswa, dan rasio dosen tetap/jumlah mahasiswa yang lebih dari rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai positif dari ke dua *variable* tersebut yang terdapat pada table *final cluster centers*. Sehingga dapat disimpulkan bahwa kelompok-2 merupakan pengelompokan dari perguruan tinggi swasta yang kurang berkualitas.

**c. Kelompok-3 (Cluster – 3)**

Karakteristik perguruan tinggi yang mengelompok pada *cluster-3* adalah hampir keseluruhan instrument penilai berada pada posisi melebihi rata-rata populasi perguruan tinggi yang diteliti, namun memiliki rasio dosen tetap/jumlah mahasiswa dibawah rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai negatif yang terdapat pada table *final cluster centers*. Sehingga dapat disimpulkan bahwa kelompok-3 merupakan pengelompokan dari perguruan tinggi yang berkualitas baik.

**d. Kelompok-4 (Cluster – 4)**

Karakteristik perguruan tinggi yang mengelompok pada *cluster-4* adalah keseluruhan instrument penilai berada pada posisi dibawah rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai negatif yang terdapat pada table *final cluster centers* dalam keseluruhan *variable*. Sehingga dapat disimpulkan bahwa kelompok-4 merupakan pengelompokan dari perguruan tinggi yang tidak berkualitas

**e. Kelompok-5 (Cluster – 5)**

Karakteristik perguruan tinggi yang mengelompok pada *cluster-5* adalah hampir keseluruhan instrument penilai berada pada posisi melebihi rata-rata populasi perguruan tinggi yang diteliti, namun memiliki rasio dosen tetap/jumlah mahasiswa dibawah rata-rata populasi perguruan tinggi yang diteliti. Hal ini terbukti dari nilai negatif yang terdapat pada table *final cluster centers*. Sehingga dapat disimpulkan bahwa kelompok-5 merupakan pengelompokan dari perguruan tinggi yang cukup berkualitas perguruan tinggi yang cukup

berkualitas.

Jumlah anggota disetiap kelompok (*cluster*) adalah untuk mengetahui jumlah anggota masing-masing kelompok yang terbentuk dapat dilihat pada tabel 3.4.

Tabel IV. Jumlah Perguruan Tinggi Swasta pada setiap Kelompok

Cluster	1	1.000
	2	1.000
	3	1.000
	4	335.000
	5	14.000
Valid		352.000

Berdasarkan tabel 5.6 dapat diketahui banyaknya anggota tiap kelompok tidak merata, dapat dilihat bahwa jumlah responden terbanyak berada di kelompok 4 ada 335 anggota, dan kelompok 5 ada 14 anggota, dan yang paling sedikit berada di kelompok 1, kelompok 2 dan kelompok 3 masing-masing ada 1 anggota. Hasil analisis terhadap masing-masing kelompok dapat diketahui bahwa Perguruan Tinggi Swasta (PTS) yang berada pada kelompok 2 dan 4 tidak berkualitas, maka disarankan perguruan tinggi yang termasuk kelompok tersebut harus dikembangkan dengan baik agar lulusannya dapat berkualitas. Adapun Perguruan Tinggi Swasta (PTS) yang berada pada kelompok 1, 3 dan 5 memiliki kualitas baik, disarankan kualitas perguruan tinggi harus lebih ditingkatkan lagi sehingga lulusan perguruan tinggi tersebut dapat memenuhi kebutuhan industri atau lapangan pekerjaan. Adapun penamaan dari hasil analisis terhadap masing-masing *cluster* dapat dilihat pada Tabel V.

Tabel V. Label Perguruan Tinggi Swasta pada Setiap Kelompok

#### 5. Evaluasi

Berdasarkan hasil evaluasi diperoleh bahwa pemodelan data mining, sudah memenuhi tujuan penelitian seperti yang telah ditentukan pada tahap pemahaman bisnis yakni pada Fase Pemahaman Bisnis (*Business Understanding Phase*).

#### 6. Penyebaran

Fase penyebaran merupakan tahap pembuatan laporan hasil penelitian, pembuatan laporan hasil dapat dilakukan setelah selesai evaluasi model pengelompokan (*clustering*).

### IV. KESIMPULAN

#### A. Kesimpulan

Dari penelitian Analisis Dan Implementasi Framework Crisp-Dm (*Cross Industry Standard Process For Data Mining*) Untuk *Clustering* Perguruan Tinggi Swasta di Kopertis IV Wilayah Jawa Barat dan Banten ini, dan berdasarkan dari analisis hasil penelitian sebelumnya, maka penulis mengambil beberapa kesimpulan, sebagai berikut :

1. *Clustering* Perguruan Tinggi untuk mengetahui ranking Perguruan Tinggi Swasta di Kopertis IV Wilayah Jawa Barat dan Banten, dengan harapan agar lebih tepat dalam melakukan

strategi, pengawasan dan peningkatkan kualitas Perguruan Tinggi yang ada di wilayah Kopertis IV tersebut.

2. Metodologi yang digunakan dalam penelitian ini adalah metodologi *CRISP-DM* (*Cross Industry Standard Process for Data Mining*), dari mulai metodologi penelitian sampai analisis hasil penelitian. Metodologi *CRISP-DM* meliputi enam fase yang harus dilakukan yaitu: *Business Understanding, Data- Understanding, Data Preparation, Modeling, Evaluation dan Devloymet*.

3. Pemodelan *Clustering* Perguruan Tinggi Swasta ini dengan menggunakan *software* SPSS Statistics Version 21 yang terdiri dari beberapa tahapan antara lain:

a. Data keseluruhan Perguruan Tinggi Kopertis IV diperoleh dari hasil transporansi ke dalam aplikasi Microsoft Office Excel 2007

b. Sebelum melakukan pengolahan data menggunakan metode *K-Means cluster*, maka dilakukan transformasi atau standarisasi data.

c. Standarisasi data dapat dilakukan dengan transformasi terhadap variabel yang relevan ke dalam bentuk *Z- Score*. Selanjutnya hasil *Z-Score* inilah yang akan dipakai dasar analisis *cluster*.

d. Melakukan pemodelan *clustering* menggunakan algoritma *K-Means*, dengan menentukan nilai *K- nya* adalah 5 (lima), yang berarti akan dibentuknya 5 *cluster* agar mendapatkan hasil *clustering* yang optimal serta analisis karakteristik Perguruan Tinggi.

e. Mengeksekusi algoritma *K-Means Clustering* agar menghasilkan *cluster* pemeringkatan Perguruan Tinggi.

f. Melakukan analisis hasil penelitian untuk memperoleh klasifikasi/*cluster* Pemeringkatan Perguruan Tinggi.

4. Model yang dihasilkan dapat dimanfaatkan sebagai acuan klasifikasi Pemeringkatan Perguruan Tinggi di

Kelompok	Label
Kelompok - 1	Terbaik
Kelompok - 2	Kurang Berkualitas
Kelompok - 3	Berkualitas Baik
Kelompok - 4	Tidak Berkualitas
Kelompok - 5	Cukup Berkualitas

Kopertis IV Wilayah Jawa Barat dan Banten. Dan dapat digunakan sebagai dasar kebijakan pengembangan Perguruan Tinggi Swasta lebih lanjut.

#### B. Saran

Dengan melihat hasil penelitian Analisis Dan Implementasi Framework Crisp-Dm (*Cross Industry Standard Process For Data Mining*) Untuk *Clustering* Perguruan Tinggi Swasta di Kopertis IV Wilayah Jawa Barat dan Banten ini, maka penulis menyarankan beberapa hal sebagai berikut :

1. Melakukan klasifikasi Perguruan Tinggi untuk

melakukan perbaikan kualitas pendidikan terus menerus dan kesehatan organisasi.

2. Melengkapi ketersediaan data baik data di PDPT maupun kelengkapan data dari Perguruan Tinggi, harus sudah disiapkan untuk dapat digunakan dalam mempermudah klasifikasi dan Pemingkatan Perguruan Tinggi

3. Penambahan variabel lainnya, agar klasifikasi dan Pemingkatan Perguruan Tinggi lebih maksimal hasilnya.

4. Hasil *Clustering* Perguruan Tinggi dapat Memberikan Informasi sebagai landasan keputusan eksekutif dalam mempromosikan kesehatan organisasi dan peningkatan kualitas Perguruan Tinggi secara terus menerus di Wilayah Kopertis IV Jawa Barat dan Banten..

5. Menjadi acuan bagi Perguruan Tinggi swasta agar dapat meningkatkan kualitas pendidikan.

6. Variasi dalam teknik clustering yaitu dengan menggunakan selain algoritma *K-Means*.

### REFERENSI

- Budiman, irwan (2012). *Data Clustering Menggunakan Metodologi CRISP-DM untuk Pengenalan Pola* Proporsi Tridharma. Yogyakarta. TESIS. Hal: 33.
- Jain, dkk ( 1999). *Data Clustering : A Review*. ACM Computing Surveys, Vol. 31, No. 3.
- Kevin E.Voges, dkk (2012). *Rough Clustering Using an Evolutionary Algorithm*. Journal of Hawaii International Conference on System Sciences.
- Larose, Daniel (2006). *Discovering Knowledge in Data: An Introduction to Data mining.*: Jhon Wiley & Son, Inc., USA.
- Mingoti, dkk (2006), “*Comparing SOM neural network with Fuzzy c-means, K-means, and traditional hierachical clustering algorithms*”, European Journal of Operational Research, Vol. 174, hal. 1742-1759
- Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia No. 2 Tahun 2016
- Peraturan Pemerintah Republik Indonesia No.4 tahun 2014 tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi.
- Sangkar. (2011), *Costumer Data Clustering Using Data maining Technique*. International Journal of Database Management System (IJDMS) Vol.3, No.4.
- Sumathi, S dkk ( 2006). *Introduction to Data Mining and its Applications*. Spinger, Verlag Berlin Heidelberg.
- Susanto, Sani dkk (2010). *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Andi, Yogyakarta.
- Steiner, J.E. (2009). *Word university Rankinga A Principal*. [www.ieausp.br/engslih/ articles](http://www.ieausp.br/engslih/articles).
- Tan, dkk (2006). *Introduction to Data Mining*. Pearson Education, Inc.
- Thomas, Binu dkk (2011). *A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining*. IInternational Journal of Recent Trends in Engineering, Vol. 1, No. 2.

Tsiptsis, dkk, (2009), *Data Mining Techniques in CRM: Inside Customer Segmentation*, John Wiley & Sons, Inc., West Sussex, U.K.

Wahono, R. S (2007). *Teknik Perangkingan PerguruanTinggi ala Webometrics*. Romi Satria Wahono. Net. <http://ban-pt.kemdiknas.go.id/dirapt.php> diakses tanggal 07 Pebruari 2016 , jam 13.00 WIB

<http://forlap.dikti.go.id/peguruantinggi> diakses tanggal 07 Pebruari 2016, jam 13.00 WIB

<http://www.kopertis4.or.id/> diakses tanggal 07 Pebruari 2016, jam 13.00 WIB

<http://www.ban-pt-universitas.com/2016/01/peringkat-ptn-pts-dijawa-barat-by-ban-pt-akreditasi.html> diakses tanggal 09 Pebruari 2016, jam 13.00 WIB

<http://simlitabmas.dikti.go.id/> diakses tanggal 08 Penruari 2016, jam 13.00 WIB

<http://www.polsri.ac.id/> pedoman- polsri/buku – himpunan - peraturan-tentang-perguruan-tinggi-di-indonesia. diakses tanggal 13 April 2016, jam 09.00 WIB