

# **PENERAPAN DATA MINING UNTUK CLUSTERING MAHASISWA RAWAN DROP OUT DENGAN METODE K-MEANS STUDI KASUS FAKULTAS ISIPOL UNIVERSITAS MUHAMMADIYAH YOGYAKARTA**

Ferdiansyah Agung Saputra, Eko Prasetyo, Asroni

*Jurusan Teknik Informatika Universitas Muhammadiyah Yogyakarta  
ferdiansyahsaputra1@gmail.com*

## **INTISARI**

*Universitas Muhammadiyah Yogyakarta merupakan perguruan tinggi swasta yang berada di Yogyakarta. UMY memiliki jumlah mahasiswa yang sangat banyak setiap Fakultasnya salah satunya Fakultas Isipol. Dengan jumlah mahasiswa yang banyak tidak menjamin semua mahasiswa tidak memiliki masalah terhadap nilai per Semester dan jumlah Sks, mahasiswa yang memiliki nilai per Semester dan Jumlah Sksnya tidak sesuai kebijakan Universitas bisa dilakukan tindak tegas atau biasa disebut dengan Drop Out. Untuk mengetahui informasi data mahasiswa Fakultas Isipol maka perlu dilakukan pengolahan data. Proses pengolahan data biasa disebut dengan Data Mining. Penelitian ini bertujuan untuk mengelompokkan mahasiswa yang bermasalah atau mahasiswa rawan Drop Out dengan menggunakan teknik Clustering. Metode yang digunakan adalah metode K-Means dengan metode ini data yang didapatkan dari data sumber akan dikelompokkan ke dalam beberapa cluster, dalam cluster terdapat data yang memiliki karakteristik yang sama sedangkan data yang memiliki karakteristik yang beda akan dikelompokkan dalam cluster yang berbeda. Hasil dari penelitian mahasiswa rawan Drop Out menggunakan Metode K-Means adalah mahasiswa yang memiliki nilai per Semester dan jumlah Sks tidak sesuai dengan kebijakan universitas setelah memiliki data mahasiswa yang bermasalah pihak universitas memberikan tindak tegas supaya mahasiswa yang bermasalah memiliki niat baik untuk memperbaiki diri sebelum dilakukannya Drop Out.*

**Kata Kunci:** *Drop Out, K-Means, Clustering, Nilai per Semester, Jumlah Sks*

## **ABSTRACT**

*Muhammadiyah University Yogyakarta is a private university located in Yogyakarta. UMY has a many students in each faculty, one of which is the Faculty of Social and Politic. With a large number of students does not secure that all students does not have problems with the value Semester and the number of credits, students who value semester and number of credits that is not with University policy can be called to as Drop Out, to know the data of students data of Faculty of Isipol need to be data processing. Data processing is usually called Data Mining. This study purpose to classify problematic students or students troubled to Drop Out by using Clustering techniques. The method used in the K-Means method with this method the data obtained from the source data will be grouped into several cluster, in the cluster there are data that have same characteristics while the data that has different characteristics will be grouped into different cluster. The results of research on Drop Out troubled students using K-Means method are students who value Semester and Number of credits is not corresponding with the University policy after having problematic student data. The university gives decisive action so that students with problems have a good intentions to improve ourselves before Drop out.*

**Keyword:** *Drop Out, K-Means, Clustering, Value Semester, Number of Credits*

## I. PENDAHULUAN

Data Mining merupakan suatu konsep yang sangat berfungsi untuk menemukan informasi yang semula kita tidak ketahui atau tersembunyi. Data Mining biasa juga di sebut dengan *Knowledge discovery (mining) in databases (KDD)* yang di mana berguna untuk meringankan *user* untuk memudahkan dalam melakukan analisis. Data Mining sangat berguna jika digunakan untuk melakukan pengolahan Data yang sangat besar karena dapat memudahkan untuk memberikan informasi yang sangat akurat.

Didalam Data Mining terdapat konsep yang dinamakan *Clustering*, yang dimana Clustering merupakan proses yang berguna untuk membagi data ke dalam bentuk kelompok berdasarkan karakteristik, dan terdapat juga metode yang dinamakan Algoritma K-means yang di mana cara kerjanya yaitu mempartisi atau membagi data ke dalam sebuah *cluster-cluster* hingga data yang memiliki kesamaan akan berada pada *cluster* yang sama dan data yang tidak sama berada pada cluster yang lain.

Banyak mahasiswa yang dari awal memulai kuliah disalah satu Universitas ternama Indonesia seperti Universitas Muhammadiyah Yogyakarta tidak bisa memprediksi dirinya apakah dia bisa lulus tepat waktu, lulus tidak tepat waktu, atau bisa saja DO (*Drop Out*). DO merupakan hal tidak di inginkan oleh seluruh Mahasiswa ataupun Mahasiswi yang sedang melakukan kuliah, DO bisa terjadi karena melakukan studi yang sangat lama (melebihi batas dari pemerintah) dan bisa terjadi karena nilai IP per semester dan jumlah Sks tidak memenuhi. Dengan adanya penelitian tentang memonitoring mahasiswa tentang rawan terjadinya DO (*Drop Out*).

Algoritma K-Means bisa digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan beberapa atribut mata kuliah. Algoritma K-Means merupakan metode mengelompokkan data berdasarkan pusat cluster (centroid awal) terdekat dari data sumber. (Asroni & Adrian, 2015)

Dalam melakukan prediksi kelulusan tepat waktu mahasiswa terdapat berbagai macam permasalahan, diantaranya yaitu metode prediksi masih menggunakan prinsip kekeluargaan sehingga dirasakan kurangnya tingkat profesionalisme dalam melakukan prediksi kelulusan tepat waktu. (Saefulloh, 2014)

Dalam pemodelan K-Means, data yang memiliki Karakteristik sama akan dikelompokkan dalam kelompok yang sama dan data yang memiliki karakteristik berbeda akan masuk kelompok yang lainnya. Penerapan model K-Means dapat digunakan

untuk mengelompokkan data seperti predikat kelulusan mahasiswa yang di dapat berdasarkan jumlah beban studi yang telah ditempuh, IPK yang diperoleh mahasiswa, dan kelulusan skripsi. (Rusdiana, 2016).

Dalam penelitian ini bertujuan untuk membagi siswa menjadi 3 *cluster* atau kelompok yang menggunakan K-Means memiliki kategorinya, Tinggi, Sedang, dan Rendah. (Naik, Shaikh, Diukar, & Dessai, 2017). Penelitian dengan menggunakan cluster ini berfungsi untuk mengurangi rasio terjadinya drop out terhadap siswa dan meningkatkan kinerja siswa menjadi lebih baik. (Singh, Rani, & Sharma, 2018)

Metode yang digunakan adalah K-Means Clustering. Dengan menggunakan metode ini data-data yang telah didapatkan dapat dikelompokkan ke dalam beberapa cluster, berdasarkan kemiripan dari data-data tersebut. (Mulyati, 2015)

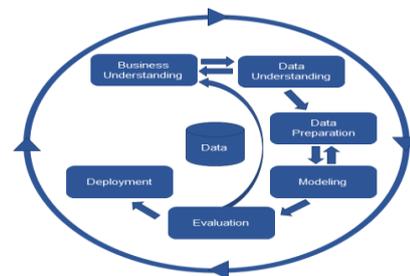
Clustering merupakan proses pembentukan suatu kelompok dari sebuah data yang tidak diketahui kelompok-kelompoknya berdasarkan kesamaan. Sehingga dalam suatu cluster terdapat kelompok yang berisi data dan karakteristik yang sama. (Wicaksono, 2016).

Data Mining merupakan ekstraksi informasi yang tersembunyi dari database yang besar. Data Mining adalah teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan agar fokus pada informasi yang paling penting dalam gudang data mereka (Ramamohan et al, 2012).

Menurut Baskoro *clustering* atau klusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam *cluster-cluster*. *Cluster* adalah sekelompok atau kumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar terhadap objek-objek yang berbeda cluster (Nango, 2012).

## II. METODOLOGI PENELITIAN

Pada penelitian rawan *Drop Out* mahasiswa Fakultas Isipol Universitas Muhammadiyah Yogyakarta menggunakan Metode CRISP-DM. Adapun proses CRISP-DM seperti pada gambar 1.



Gambar 1 Metode CRISP-DM 1

### 1. Business Understanding

*Business Understanding* atau pemahaman data adalah proses dimana kita mengerti dengan apa yang akan dilakukan dan mengetahui hasil yang diinginkan

### 2. Data Understanding

*Data Understanding* atau biasa disebut dengan Pemahaman Data adalah data yang akan diteliti sudah dipahami tentang atribut apa saja yang akan digunakan dan menemukan masalah yang akan timbul pada data.

### 3. Data Preparation

*Data Preparation* atau biasa disebut dengan Persiapan Data secara garis besar proses yang akan dilakukan adalah memperbaiki masalah pada data.

### 4. Modeling

Modeling atau biasa disebut pemodelan menentukan metode yang digunakan. Pada penelitian ini metode yang digunakan adalah K-Means Clustering. Clustering memiliki proses yang pertama menentukan centroid awal yang dilakukan secara random atau acak, proses kedua adalah menentukan jarak yang dimana memiliki rumus seperti pada gambar 2.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1, 2, 3, 4, 5, 6, 7, 8$$

Gambar 2 Rumus Menentukan Jarak

Terakhir proses clustering adalah menentukan cluster pada setiap data yang diteliti.

### 5. Evaluation

Evaluation atau evaluasi secara garis besar hasil dari pemodelan akan diperiksa kembali dan jika ada masalah akan dilakukan evaluasi.

## III. HASIL DAN PEMBAHASAN

Data yang diuji adalah data Fakultas Isipol Universitas Muhammadiyah Yogyakarta yang terdiri dari tiga

Jurusan antara lain Hubungan International (HI), Ilmu Pemerintahan (IP), dan terakhir Ilmu Komunikasi (IK) untuk mendukung penelitian ini terdapat 5 atribut yang mendukung yaitu. Nim, Id Jurusan, tahun masuk, semester, nilai persemester, dan Jumlah SKS.

1. Menentukan centroid awal dilakukan secara random atau acak. Adapun centroid awalnya bisa dilihat pada table 1.

Tabel 1 Centroid Awal

3.75	3.43	3.25	3.31	3.63	3.80	4.00	3.70
3.47	3.07	2.87	3.47	3.14	3.15	4.00	3.50
3.38	2.30	1.69	0.00	0.27	0.50	0.67	0.00

2. Menentukan jarak disetiap data yang akan dijadikan untuk menentukan cluster. Sebagai contoh mahasiswa yang memiliki Nim 20120510039 yang dimana nilai persemesternya seperti pada tabel 2.

Tabel 2 Contoh Data Mahasiswa

1	2	3	4	5	6	7	8
2.84	0.42	1.80	0.38	2.15	1.67	0.50	1.00

Dengan data diatas terdapat hasil jarak seperti berikut:

jarak 1 :  $C_0(x,y) = \sqrt{(2.84-3.75)^2+(0.42-3.43)^2+(1.80-3.25)^2+(0.38-3.31)^2+(2.15-3.63)^2+(1.67-3.80)^2+(0.50-4.00)^2+(1.00-3.70)^2} = 6.85$

jarak 2 :  $C_1(x,y) = \sqrt{(2.84-3.47)^2+(0.42-3.07)^2+(1.80-2.87)^2+(0.38-3.47)^2+(2.15-3.14)^2+(1.67-3.15)^2+(0.50-4.00)^2+(1.00-3.50)^2} = 6.31$

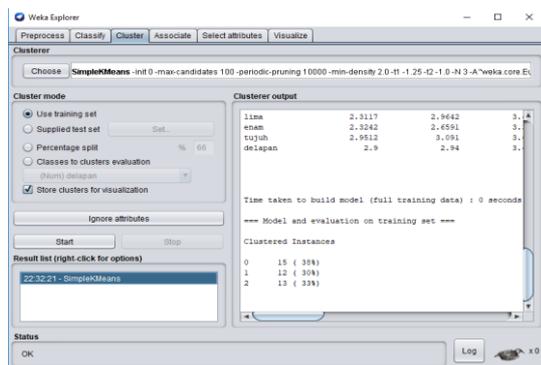
jarak 3 :  $C_2(x,y) = \sqrt{(2.84-3.38)^2+(0.42-2.30)^2+(1.80-1.69)^2+(0.38-0.00)^2+(2.15-0.27)^2+(1.67-0.50)^2+(0.50-0.67)^2+(1.00-0.00)^2} = 3.15$

3. Menentukan hasil cluster seperti pada Tabel 3.

Tabel 3 Hasil Cluster

Nim	Jarak	Jarak	Jarak	Cluster	Cluster	Cluster
	1	2	3	0	1	2
20130510426	6.85	6.31	3.15			V

Adapun hasil dari Weka bias dilihat pada gambar 3 dan grafik hasil weka bisa dilihat pada gambar 4.



Gambar 3 Hasil Dari Contoh Data Mahasiswa



Gambar 4 Hasil Grafik Contoh Data Mahasiswa

1. Terdapat 3 cluster yang dimana setiap cluster memiliki persen. Seperti cluster 0 memiliki 15 data (38%), cluster 1 memiliki 12 data (30%), dan cluster 2 memiliki 13 data (33%).
2. Memiliki rata-rata nilai disetiap cluster, cluster 0 memiliki rata-rata nilai 3.25, cluster 1 memiliki rata-rata nilai 3.48, sedangkan cluster 2 memiliki rata-rata nilai 0.8 yang artinya cluster 2 merupakan sekumpulan mahasiswa yang memiliki masalah terhadap nilai semesternya.

#### IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan tentang Mahasiswa Rawan Drop Out Fakultas Isipol Universitas Muhammadiyah Yogyakarta disimpulkan bahwa Metode K-Means bisa digunakan untuk melakukan pengelompokan atau cluster Mahasiswa berdasarkan nilai persemesternya.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penelitian Mahasiswa Rawan Drop Out Fakultas Isipol Universitas Muhammadiyah Yogyakarta.

#### REFERENSI

- Asroni, & Adrian, R. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang.
- Mulyati, S. (2015). Penerapan Data Mining Dengan Metode Clustering Untuk Pengelompokan Data Pengiriman Burung.
- Naik, P., Shaikh, R., Diukar, O., Dessai, S., & Project Guide], P. S. B. (2017). Predicting Student Performance Based On Clustering And Classification.
- Rusdiana, L. (2016). Pemodelan K-Means Pada Penentuan Predikat Kelulusan Mahasiswa STMIK Palangkaraya, 6.
- Saefulloh, A. (2014). Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu.
- Singh, M., Rani, A., & Sharma, R. (2018). An Optimised Approach for Student's Academic Performance by K-Means Clustering Algorithm Using Weka Interface, 2(7), 5.
- Wicaksono, A. (2016). Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMAN 16 Bekasi).