

# Analisis Pola Aktivitas Mahasiswa Lulus Tepat Waktu berbasis *Data Mining* dan *Process Mining* (Studi Kasus Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya)

Hafizh Yuwan Fauzan<sup>1</sup>, Nanang Yudi Setiawan<sup>2</sup>, Ahmad Afif Supianto<sup>3</sup>

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: <sup>1</sup>hayeeef8@student.ub.ac.id, <sup>2</sup>nanang@ub.ac.id, <sup>3</sup>afif.supianto@ub.ac.id

## Abstrak

Permasalahan program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya adalah adanya beberapa mahasiswa yang masih belum lulus walaupun sudah melewati semester 8. Hal tersebut dapat mempengaruhi penilaian reakreditasi, karena jika mahasiswa sampai drop out dikarenakan mencapai batas semester yang telah ditentukan maka akan berpengaruh pada penilaian reakreditasi program studi Sistem Informasi. Untuk itu diperlukan langkah awal yang dapat membuat mahasiswa lulus tepat waktu. Solusi yang ditawarkan pada penelitian ini adalah dengan membuat klaster berdasarkan atribut akademik. Setelah itu dilakukan *Process Mining* pada tiap klaster agar bisa didapatkan informasi dan hubungan yang tidak terlihat pada data pengerjaan skripsi. Penelitian ini akan menggunakan algoritma K – Means++ untuk *Data Mining* dan algoritma Fuzzy Miner untuk *Process Mining*. Hasil dari proses klasterisasi nya didapatkan empat klaster dengan karakteristiknya masing – masing, dan juga pada tiap klaster tersebut dapat dilihat informasi statistik dan model proses pengerjaan skripsinya.

**Kata kunci:** *Data Mining, Process Mining, k-means++, fuzzy miner, Silhouette Coefficient*

## Abstract

*One of many problems at Information System at Faculty of Computer Science Universitas Brawijaya is that there is some undergraduate who have not graduated even though they have passed 8th semester. This can affect the accreditation assessment, because if students drop out due to reaching the semester limit then it will affect the accreditation assessment. For this very reason, initial steps are needed that can make students graduate on time. The solution offered in this study is to create a cluster, based on academic attributes. After that, Process Mining is performed in each cluster so that information and relationships that cannot be obtained in the data. This research will use K-Means ++ algorithm for Data Mining and Fuzzy Miner algorithm for Process Mining. The results of the clustering process are four clusters obtained with their respective characteristics, and also for each of these clusters can be seen statistical information and models process of the thesis.*

**Keywords:** *Data Mining, Process Mining, k-means++, fuzzy miner, Silhouette Coefficient*

## 1. PENDAHULUAN

Universitas Brawijaya khususnya Fakultas Ilmu Komputer merupakan salah satu instansi perguruan tinggi yang mempunyai akreditasi di tiap program studinya, salah satunya adalah program studi Sistem Informasi dimana mendapat akreditasi A. Kepala Program Studi Yusi Tyroni Mursityo, S.Kom., M.AB. mengatakan salah satu permasalahan program studi Sistem Informasi adalah adanya beberapa

mahasiswa yang masih belum lulus walaupun sudah melewati semester 8. Hal tersebut dapat mempengaruhi penilaian reakreditasi, karena jika mahasiswa sampai drop out dikarenakan mencapai batas semester yang telah ditentukan maka akan berpengaruh pada penilaian reakreditasi program studi Sistem Informasi. Untuk itu diperlukan langkah awal yang dapat membuat mahasiswa lulus tepat waktu.

Permasalahan seperti ini tidak hanya terjadi di prodi Sistem Informasi sehingga sudah ada

beberapa penelitian yang berhubungan dengan kelulusan mahasiswa, contohnya adalah penelitian yang dilakukan oleh Alvi Syahrin (2013) yang berjudul “Implementasi Algoritma K-Means Untuk Klasterisasi Mahasiswa Berdasarkan Prediksi Waktu Kelulusan,” pada penelitian tersebut peneliti mencoba untuk mengklasterisasi mahasiswa berdasarkan data akademik yang kemudian dapat membantu pihak universitas untuk memprediksi waktu kelulusan golongan mahasiswa. Pada penelitian tersebut algoritma K – Means digunakan untuk mengklasterisasi mahasiswa dengan atribut akademik. Pada penelitian ini algoritma K – Means juga digunakan untuk mengklasterisasi mahasiswa, akan tetapi terdapat juga tambahan *Process Mining* agar bisa mendapatkan informasi yang belum ada pada penelitian sebelumnya.

Selain penelitian tentang *Data Mining* terdapat juga penelitian tentang *Process Mining* pada sistem edukasi. Artikel ilmiah internasional yang berjudul “*Process Mining* Online Assessment Data,” menjelaskan tentang bagaimana potensi dari *Process Mining* dalam mendapatkan informasi dan hubungan yang tidak terlihat dari suatu proses dalam edukasi. Peneliti pada penelitian tersebut menggunakan Tool ProM untuk mendapatkan informasi yang tidak bisa didapatkan oleh *Data Mining* konvensional yang berfokus pada data, bukan pada proses (Pechenizkiy, Trčka, Vasilyeva, Van Der Aalst, & De Bra, 2009). Saat ini terdapat dua pendekatan fuzzy miner terpercaya, yaitu pendekatan Fuzzy Miner dengan menggunakan ProM dan juga pendekatan Fuzzy Miner dengan menggunakan Disco Fluxicon (Premchaiswadi & Porouhan, 2015). Dikarenakan pada penelitian oleh Pechenizkiy dan kawan – kawan sudah menggunakan ProM, oleh karena itu pada penelitian kali ini menggunakan Tool dari Disco.

Manfaat adanya penelitian ini adalah pemegang keputusan di Program Studi Sistem Informasi dapat melihat pola pengerjaan skripsi mahasiswa di tiap klaster. Dengan diketahuinya pola suatu mahasiswa maka pemegang keputusan di Program Studi Sistem Informasi mendapatkan informasi dan hubungan yang tidak terlihat pada data mentah sebelumnya. Dari informasi tersebut pemegang keputusan dapat melihat klaster mana yang membutuhkan perhatian yang lebih. Informasi tersebut berupa grafik kelulusan, frekuensi pengerjaan tiap aktivitas, durasi pengerjaan tiap aktivitas, durasi

rata – rata tiap aktivitas, durasi terlama dan tercepat tiap aktivitas dan sebagainya. Dan juga setelah dianalisis maka pemegang keputusan dapat memprioritaskan untuk membantu mahasiswa yang masuk ke klaster terburuk dalam pengerjaan skripsinya.

Berdasarkan permasalahan yang telah dijelaskan sebelumnya, penelitian ini akan menggunakan algoritma K-Means++ untuk melakukan klasterisasi terhadap mahasiswa pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. Alasan mengapa menggunakan variasi algoritma K – Means dibandingkan klaster hirarki adalah karena pada algoritma klaster hirarki lebih cocok untuk data yang bertipe kategori sedangkan data yang digunakan pada penelitian ini adalah numerik (Kaushik & Mathur, 2014). Selain itu juga tidak terdapat perubahan performa yang signifikan antara algoritma klasterisasi unsupervised lainnya. Dan penggunaan algoritma ini lebih baik dalam dataset yang berukuran besar yaitu dengan atribut diatas 20 (Nathan & Scobell, 2012). Variasi K – Means++ digunakan karena terdapat perbaikan performa dibandingkan K – Means tanpa modifikasi. Perbedaan mendasar dari K – Means++ dan K – Means pada umumnya terdapat pada pemilihan titik awal centroid. Jadi K – Means++ dapat memberikan kecepatan dan hasil yang lebih baik daripada K – Means pada umumnya (Agarwal, Yadav, & Singh, 2012).

Setelah mahasiswa terpisah menurut klaster maka akan dilakukan *Process Mining* dengan Disco untuk mencari informasi yang tidak terlihat dalam proses pengerjaan skripsi yang dilakukan mahasiswa pada tiap klaster. Dan setelah itu akan dilakukan analisis perbandingan tiap klaster untuk mengetahui apakah ada pengaruh antara aktivitas akademik dengan proses pengerjaan skripsi yang dapat dijadikan bahan pertimbangan bagi pemegang keputusan program studi dalam membuat kebijakan. Oleh karena itu penelitian ini berjudul “Analisis Pola Aktivitas Mahasiswa Lulus Tepat Waktu Bebas *Data Mining* Dan *Process Mining* (Studi Kasus Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya)”.

## 2. LANDASAN KEPUSTAKAAN

Menurut Turban (Turban, Aronson, & Liang, 2005), *Data Mining* adalah proses yang menggunakan statistik, matematika, kecerdasan buatan dan machine learning untuk

mengekstraksi dan mengidentifikasi informasi yang berguna dan informasi dari big database. Sebelum pengertian tersebut, istilah ini digunakan untuk menggambarkan proses identifikasi pola-pola data. Namun, seiring berkembangnya zaman, definisi awal diubah untuk memasukkan sebagian besar jenis analisis data.

*Process Mining* dapat dibidang memiliki tujuan untuk mencari, mengawasi, dan meningkatkan fungsi proses yang sudah ada. Oleh sebab itu *Process Mining* merupakan sebuah Teknik, yang dapat digunakan untuk mengumpulkan informasi dari log aktivitas. (Tadanai & Tanuska, 2015).

*Process Mining* merupakan sebuah teknik yang dapat memproses pengumpulan informasi dari suatu *Event Logs*. *Event Logs* merupakan tempat dimana aktivitas dari sistem berada, tiap event log berisi tentang informasi yang sudah di proses oleh sistem IT pada suatu waktu. (Tadanai & Tanuska, 2015)

Teknik – teknik pada *Data Mining* sendiri ada beberapa, yaitu:

- Classification: menyimpulkan karakteristik penentu dari suatu kelompok tertentu. Metode-metode ini melibatkan penyamaan seperangkat data dengan seperangkat segmen yang sudah diketahui, dan memetakan semua data lain (pelanggan) ke dalam set ini. Decision trees dan neural networks adalah teknik yang sering digunakan.

- Clustering: mengidentifikasi kelompok data yang memiliki karakteristik yang mirip. Clustering melakukan pendekatan pemecahan masalah segmentasi. Algoritma clustering dapat digunakan untuk melakukan identifikasi terhadap pelanggan dengan kebutuhan tertentu.

- Association: mengidentifikasi hubungan antar kejadian yang terjadi setiap saat. Pendekatan asosiasi menangani masalah yang ditandai dengan analisis market basket. Dalam ritel, ada upaya untuk mengidentifikasi produk apa yang dijual dengan produk yang lain, dan sampai tingkat apa. Metode statistik biasanya metode yang sering digunakan.

- Sequencing: mirip dengan pendekatan association, perbedaannya adalah hubungan terjadi selama periode waktu tertentu.

- Regression: digunakan untuk memetakan data ke nilai prediksi. Teknik linier dan nonlinier digunakan. Teknik ini berupa estimasi, sering melibatkan identifikasi metrik dan mengevaluasi data (kustomer) Di sepanjang metrik dengan menetapkan skor. Prediksi

penjualan dapat dicapai juga.

- Forecasting: memperkirakan nilai masa depan berdasarkan pola dalam kumpulan data yang besar. Teknik ini adalah bentuk lain dari prakiraan. Forecasting adalah upaya untuk memanfaatkan metode deret waktu statistik untuk memprediksi penjualan di masa depan.

Perbedaan mendasar dari K – Means++ dan K – Means pada umumnya terdapat pada pemilihan titik awal centroid. Jadi K – Means++ dapat memberikan kecepatan dan hasil yang lebih baik daripada K – Means pada umumnya. Algoritma ini di kembangkan oleh David Arthur dan Sergei Vassilvitskii, sebagai algoritma yang menghindari klustering yang kurang baik pada K – Means normal. K – Means++ pada dasarnya memiliki algoritma seperti dibawah ini:

1. Memilih centeroid awal secara acak berseragam dari X
2. Memilih centeroid berikutnya
3. Ulangi langkah kedua sampai kita memiliki total sejumlah k
4. Ulangi langkah 2-4 dengan k-means normal

Algoritma Fuzzy Miner merupakan salah satu algoritma Process Discovery terbaru. Proses pada algoritma ini membuat pengguna untuk menjelajahi proses dari *Event Logs* secara interaktif. Fuzzy Miner sangat cocok digunakan untuk menggali proses yang lebih tidak terstruktur. (Vijayarani et al., 2018)

*Silhouette Coefficient* merupakan tolak ukur untuk menentukan kualitas dan kekuatan suatu klaster, objek dalam klaster dapat dilihat baik atau tidaknya.

### 3. METODOLOGI PENELITIAN



Gambar 1. Alur Metodologi Penelitian

Proses identifikasi masalah dilakukan dengan melakukan observasi dan wawancara pada bagian akademik dan Kepala Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya.

Studi literatur diperlukan sebagai dasar teori yang digunakan untuk membantu penulisan penelitian skripsi dan berkaitan dengan metode yang digunakan dalam penelitian tentang klusterisasi menggunakan algoritma K-Means. Sumber yang biasa digunakan untuk mendapatkan keterangan serta penjelasan mengenai teori-teori yang relevan dengan penelitian yang dilakukan antara lain mengenai *Data Mining*, K-Means, database, pengukuran hasil klusterisasi, dan pengujian Usability.

Pada penelitian ini data primer yang didapatkan berasal dari bagian akademik Fakultas Ilmu Komputer Universitas Brawijaya. Ada dua macam data yang di ambil yaitu data akademik mahasiswa dan data tanggal pengerjaan skripsi mahasiswa. Data akademik mahasiswa diambil dari bagian akademik Program Studi Sistem Informasi, sedangkan data tentang tanggal pengerjaan skripsi diambil dari bagian PSIK FILKOM.

Setelah data didapatkan maka tahap selanjutnya adalah pembersihan data. Pembersihan data dilakukan dengan menghilangkan data yang tidak relevan dan tidak konsisten. Setelah itu akan dilakukan proses seleksi data. Karena data yang diminta banyak maka hanya data – data dan atribut yang

dibutuhkan saja yang diambil.

Setelah Preprocessing data selesai maka tahapan selanjutnya adalah untuk mengimplementasikan *Data Mining*. *Data Mining* digunakan untuk mendapatkan kluster – kluster yang akan digunakan pada tahap selanjutnya yaitu *Process Mining*.

Data akademik yang telah melalui Preprocessing akan dikelompokkan dengan menggunakan metode K – Means++ Clustering. Pada metode K – Means++ terdapat dua tahapan yaitu, menentukan nilai K (jumlah kluster) dengan menggunakan metode *Silhouette Coefficient*, dan proses Clustering itu sendiri.

Setelah dilakukan klusterisasi, maka akan diimplementasikan teknik *Process Mining* untuk mengetahui urutan aktivitas yang dilakukan oleh tiap cluster. Proses ini akan menggunakan *Disco Tools*. *Disco* sendiri menggunakan algoritma fuzzy yang sudah di modifikasi. Hal ini karena pada *Disco Tools* terdapat beberapa statistik yang dapat berguna bagi mahasiswa seperti rata – rata durasi pengerjaan skripsi, durasi tersingkat pengerjaan skripsi, durasi terlama pengerjaan skripsi, frekuensi dari pengerjaan skripsi dan sebagainya.

Setelah tahapan dari *Data Mining* dan *Process Mining* selesai maka tahapan selanjutnya adalah menganalisis hasil dari kedua tahapan tersebut. Pada tahapan ini tiap kluster akan di analisis karakteristik nya dan juga perbedaan yang mencolok dari kluster lainnya. Selain analisis hasil *Data Mining*, tahapan *Process Mining* juga akan di analisis. Pada analisis bagian *Process Mining* akan dilihat bagaimana karakteristik dari proses pengerjaan skripsi tiap kluster. Selain itu juga akan dilakukan analisis pada bagian model proses nya. Setelah kedua tahapan selesai di analisis maka akan terlihat kluster mana yang paling optimal dan juga kluster mana yang butuh perhatian yang khusus dari pemegang keputusan program studi.

Tahap terakhir dari penelitian skripsi ini adalah pengambilan kesimpulan dan saran. Kesimpulan diambil setelah semua tahap-tahap mulai dari Preprocessing, implementasi *Data Mining* dengan K-Means, implementasi *Process Mining* dengan *Disco* dan sampai pada hasil *Data Mining* dan *Process Mining* telah selesai dilakukan. Saran berguna untuk memperbaiki

kesalahan dan diharapkan dapat menyempurnakan penulisan serta untuk memberikan pertimbangan untuk penelitian selanjutnya.

**4. PENGUMPULAN DATA DAN PREPROCESSING**

Pada penelitian ini data yang dibutuhkan adalah data akademik dan juga data tanggal – tanggal pengerjaan skripsi mahasiswa. Data akademik didapatkan dari bagian akademik program studi sistem informasi. Data yang diambil memiliki format .xlsx yang berjumlah 380 rows dan 55 coloumn. Data tersebut diberikan dalam bentuk softfile melalui flashdisk. Sedangkan untuk data tanggal – tanggal pengerjaan skripsi didapatkan dari PSIK FILKOM UB. Data tersebut juga diberikan dalam format .xlsx dan memiliki jumlah mahasiswa yang sama dengan data dari akademik program studi sistem informasi. Oleh karena itu data yang didapatkan dalam penelitian ini adalah data sekunder karena diperoleh dari database Fakultas Ilmu Komputer Universitas Brawijaya. Data yang didapatkan dapat dilihat pada tabel 1.

Selain data akademik dan pengerjaan skripsi, pada penelitian ini terdapat juga wawancara yang dilakukan oleh kepala program studi Sistem Informasi FILKOM UB. Data tersebut digunakan untuk mendapatkan permasalahan di program studi yang kemudian dimuat pada latar belakang penelitian. Untuk pertanyaan seputar wawancara terdapat di lampiran.

Tabel 1. Atribut yang Dipakai dalam Penelitian

No	Nama Atribut	Keterangan
1	ID MHS	ID untuk menggantikan NIM mahasiswa
2	Angkatan	Tahun Angkatan mahasiswa
3	IPK	IPK mahasiswa setelah lulus
4	IPL SMT 1	IP lulus mahasiswa ketika semester 1
5	IPB SMT 1	IP beban mahasiswa ketika semester 1
6	SKSL SMT 1	SKS lulus mahasiswa ketika semester 1
7	SKSB SMT 1	SKS beban mahasiswa ketika semester 1

8	IPL SMT 2	IP lulus mahasiswa ketika semester 2
9	IPB SMT 2	IP beban mahasiswa ketika semester 2
10	SKSL SMT 2	SKS lulus mahasiswa ketika semester 2
11	SKSB SMT 2	SKS beban mahasiswa ketika semester 2
12	IPL SMT 3	IP lulus mahasiswa ketika semester 3
13	IPB SMT 3	IP beban mahasiswa ketika semester 3
14	SKSL SMT 3	SKS lulus mahasiswa ketika semester 3
15	SKSB SMT 3	SKS beban mahasiswa ketika semester 3
16	IPL SMT 4	IP lulus mahasiswa ketika semester 4
17	IPB SMT 4	IP beban mahasiswa ketika semester 4
18	SKSL SMT 4	SKS lulus mahasiswa ketika semester 4
19	SKSB SMT 4	SKS beban mahasiswa ketika semester 4
20	IPL SMT 5	IP lulus mahasiswa ketika semester 5
21	IPB SMT 5	IP beban mahasiswa ketika semester 5
22	SKSL SMT 5	SKS lulus mahasiswa ketika semester 5
23	SKSB SMT 5	SKS beban mahasiswa ketika semester 5
24	IPL SMT 6	IP lulus mahasiswa ketika semester 6
25	IPB SMT 6	IP beban mahasiswa ketika semester 6
26	SKSL SMT 6	SKS lulus mahasiswa ketika semester 6
27	SKSB SMT 6	SKS beban mahasiswa ketika semester 6
28	IPL SMT 7	IP lulus mahasiswa ketika semester 7
29	IPB SMT 7	IP beban mahasiswa ketika semester 7
30	SKSL SMT 7	SKS lulus mahasiswa ketika semester 7
31	SKSB SMT 7	SKS beban mahasiswa ketika semester 7
32	IPL SMT 8	IP lulus mahasiswa ketika semester 8

33	IPB SMT 8	IP beban mahasiswa ketika semester 8
34	SKSL SMT 8	SKS lulus mahasiswa ketika semester 8
35	SKSB SMT 8	SKS beban mahasiswa ketika semester 8
36	IPL SMT 9	IP lulus mahasiswa ketika semester 9
37	IPB SMT 9	IP beban mahasiswa ketika semester 9
38	SKSL SMT 9	SKS lulus mahasiswa ketika semester 9
39	SKSB SMT 9	SKS beban mahasiswa ketika semester 9
40	IPL SMT 10	IP lulus mahasiswa ketika semester 10
41	IPB SMT 10	IP beban mahasiswa ketika semester 10
42	SKSL SMT 10	SKS lulus mahasiswa ketika semester 10
43	SKSB SMT 10	SKS beban mahasiswa ketika semester 10
44	IPL SMT 11	IP lulus mahasiswa ketika semester 11
45	IPB SMT 11	IP beban mahasiswa ketika semester 11
46	SKSL SMT 11	SKS lulus mahasiswa ketika semester 11
47	SKSB SMT 11	SKS beban mahasiswa ketika semester 11
48	Tgl pengajuan proposal	Tanggal mahasiswa mengajukan proposal skripsi
49	Tgl verifikasi akademik	Tanggal mahasiswa diverifikasi oleh akademik dalam tahap pengerjaan skripsi
50	Tgl penentuan calon pembimbing	Tanggal mahasiswa ditentukan dosen pembimbingnya
51	Tgl penerbitan st	Tanggal penerbitan surat tugas skripsi bagi mahasiswa
52	Tgl P0	Tanggal mahasiswa melakukan proses P0
53	Tgl P1	Tanggal mahasiswa melakukan proses P1
54	Tgl P2	Tanggal mahasiswa melakukan proses P2
55	Tgl semhas	Tanggal mahasiswa melaksanakan seminar hasil skripsi

56	Tgl ujian	Tanggal mahasiswa melaksanakan ujian skripsi
----	-----------	--

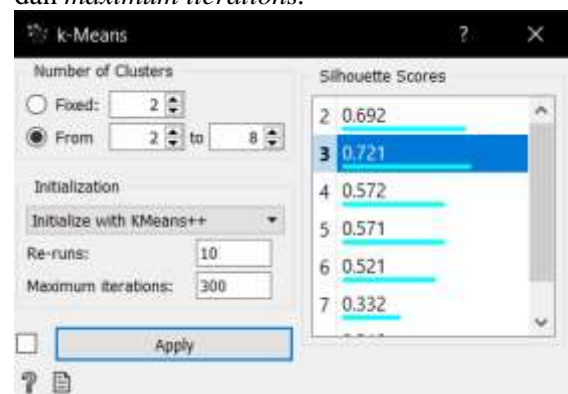
Pembersihan data diperlukan karena masih ada beberapa data yang bersifat noise maupun tidak relevan dengan analisa data yang akan dilakukan. Contohnya pembersihan pada atribut yang belum memuat nilai numerik, secara spesifik yaitu pada atribut yang digunakan sebagai fitur data dalam tahapan *Data Mining* seperti fitur IP yang masih bernilai N/A.

Data yang noise dan tidak relvan akan mempengaruhi model pengelompokan yang terbentuk dari tahapan *Data Mining*. Sedangkan pada analisa aktivitas dengan *Process Mining*, pembersihan data khususnya pada data jadwal pengerjaan skripsi akan memvalidasi tanggal pelaksanaan setiap aktivitas dalam rentang waktu yang telah ditetapkan oleh program studi pada semester ganjil dan genap setiap tahunnya.

### 5. IMPLEMENTASI

Pada tahap ini data kinerja akademik mahasiswa akan diproses menggunakan *Data Mining* sehingga siap untuk dilanjutkan ke tahap selanjutnya. Atribut yang dipakai pada tahap ini sudah dibahas sebelumnya pada tabel 1.

Jumlah *Re-runs* dan *Maximum iterations* pilihan bawaan *Tool Orange* adalah 10 dan 300. Sebelum menentukan jumlah yang akan di pakai pada penelitian ini maka peneliti membandingkan terlebih dahulu apakah terdapat perbedaan yang signifikan antar jumlah *re-runs* dan *maximum iterations*.



Gambar 2. Konfigurasi Bawaan *Tool Disco*

Proses percobaan tersebut dilakukan secara manual sampai tidak ada perbedaan yang signifikan pada hasil dari *Silhouette Scores* nya. Setelah dilakukan percobaan untuk jumlah *Re-runs* dan *maximum iterations* dapat ditarik kesimpulan untuk jumlahnya masing – masing adalah 30 dan 300, seperti yang dapat dilihat



Attribute digunakan untuk menunjukkan aktivitas apa yang sedang dilakukan mahasiswa, dan value merupakan tanggal aktivitas tersebut dilakukan. Tabel 4 merupakan salah satu contoh untuk klaster 1 dari total empat klaster. Jumlah total row yang didapat berjumlah 64 yang sudah termasuk dari header.

Tabel 4. Hasil Data Setelah di Ubah Menjadi Horizontal

ID	Attribute	Value
MHS190	Tgl Pengajuan Proposal	9/30/2015
MHS190	Tgl Verifikasi Akademik	10/8/2015
MHS190	Tgl Penentuan Calon Pembimbing	10/9/2015
MHS190	Tgl Penerbitan ST	10/12/2015
MHS190	Tgl P0	10/26/2015
MHS190	Tgl P1	12/15/2015
MHS190	Tgl P2	4/7/2016
MHS190	Tgl Semhas	5/26/2016
MHS190	Tgl Ujian	7/21/2016
MHS191	Tgl Pengajuan Proposal	10/3/2015
MHS191	Tgl Verifikasi Akademik	10/9/2015
MHS191	Tgl Penentuan Calon Pembimbing	10/9/2015
MHS191	Tgl Penerbitan ST	10/12/2015
MHS191	Tgl P0	10/26/2015
MHS191	Tgl P1	6/3/2016
MHS191	Tgl P2	6/21/2016
MHS191	Tgl Semhas	6/24/2016
MHS191	Tgl Ujian	7/29/2016
MHS192	Tgl Pengajuan Proposal	10/3/2015
MHS192	Tgl Verifikasi Akademik	10/9/2015
MHS192	Tgl Penentuan Calon Pembimbing	10/9/2015
MHS192	Tgl Penerbitan ST	10/12/2015
MHS192	Tgl P0	3/17/2016
MHS192	Tgl P1	6/20/2016
MHS192	Tgl P2	6/23/2016
MHS192	Tgl Semhas	6/29/2016
MHS192	Tgl Ujian	7/29/2016
MHS193	Tgl Pengajuan Proposal	9/30/2015

Pada penelitian kali ini ID dipilih menjadi ID Disco, kemudian untuk kolom kedua dipilih menjadi Activity dan kolom terakhir menjadi timestamp. Apabila semua konfigurasi telah sesuai dengan masukan yang digunakan Tool Disco maka akan ada tulisan “Ready to start import” di kanan bawah. Setelah muncul tulisan tersebut maka di klik tombol “Start Import”.

Setelah proses pengaturan selesai Tool Disco akan memproses data kemudian membuat proses bisnis. Proses diatas akan dilakukan secara berulang – ulang sejumlah jumlah klaster yang diperlukan. Dikarenakan pada penelitian ini jumlah klaster nya sebanyak 4 klaster oleh karena itu proses tersebut di aplikasikan kedalam 4 klaster tersebut secara masing – masing. Setelah diproses maka akan muncul tiga buah jendela yang berisi informasi penting dalam Tool

Disco.

Setelah proses tersebut dilakukan akan muncul beberapa jendela pada tampilan utama tool Disco. Pada jendela tersebut terdapat beberapa informasi seperti proses bisnis nya seperti apa dan juga disebelah kanan terdapat detail dari proses bisnis tersebut. Pada bagian proses model anak panah yang lebih tebal menunjukkan bahwa proses tersebut banyak yang melewati dibandingkan dengan anak panah yang tipis.

Jendela kedua dari Tool Disco merupakan jendela statistik. Pada jendela ini terdapat berbagai macam informasi statistik. Dimulai dari informasi yang paling kanan terdapat jumlah events, jumlah kasus, jumlah aktivitas, durasi median kasus, durasi mean kasus, tanggal mulai dan selesai.

Jendela terakhir pada Tool Disco merupakan jendela Cases. Informasi terpenting dari jendela ini adalah Variant dari suatu proses. Variant disini merupakan beberapa macam path yang digunakan pada suatu proses.

## 6. ANALISIS HASIL DAN KESIMPULAN

Setelah melakukan Data Mining dan juga Process Mining pada tiap mahasiswa tahapan selanjutnya adalah untuk menganalisis hasil dari proses – proses tersebut. Tahapan pertama dari proses analisis adalah dengan mencari karakteristik tiap klaster mahasiswa terlebih dahulu. Proses ini digunakan untuk dapat mengetahui perbedaan tiap klaster yang kemudian dapat dijadikan acuan untuk tahap analisis hasil Process Mining.

Berdasarkan dari hasil Data Mining yang dilakukan terhadap mahasiswa Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya angkatan 2014 dan 2015 yang sudah lulus terdapat jumlah total 372 mahasiswa yang terbagi menjadi empat klaster. Keempat klaster tersebut di dikelompokkan berdasarkan 47 atribut akademik Jumlah dari klaster 1 adalah sebanyak 7 mahasiswa, klaster 2 sebanyak 184 mahasiswa, klaster 3 sebanyak 173 mahasiswa dan klaster 4 sebanyak 8 mahasiswa. Dari semua data tersebut mahasiswa angkatan 2014 sebanyak 196 atau 52,7%, sedangkan mahasiswa angkatan 2015 sebanyak 176 mahasiswa atau 47,3%.

Berdasarkan hasil analisis dari penelitian ini dapat ditarik beberapa kesimpulan yaitu:

1. Kesimpulan pertama adalah tidak dimasukkannya atribut angkatan pada proses



*Data Mining* akan tetapi masing – masing klaster memiliki angkatan yang sama untuk anggota nya mengindikasikan terdapat perbedaan performa akademik antar angkatan. Dengan adanya indikasi ini pemegang keputusan pada program studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya dapat memperhatikan angkatan yang kesusahan dalam akademik maupun pengerjaan skripsi. Dalam penelitian kali ini angkatan yang sudah tua lah yang memiliki performa yang tidak sebaik angkatan dibawahnya.

2. Proses pencatatan aktivitas skripsi juga harus diperhatikan lagi. Dari kesimpulan yang didapatkan masih banyak tanggal aktivitas pengerjaan skripsi yang berbeda jauh dari data lainnya. Contohnya adalah ada mahasiswa yang melakukan P2 setelah ujian skripsi. Hal ini menunjukkan bahwa standar pencatatan pengerjaan skripsi masih belum tegas dalam prakteknya. Akan tetapi kesalahan – kesalahan tersebut masih dapat dibilang sedikit dibandingkan keseluruhan data, jadi tidak menyebabkan adanya kesalahan yang signifikan. Hal ini ditekankan agar apabila ada penelitian yang mirip seperti ini lagi dapat membuat penelitian tersebut lebih akurat.

3. Selain kesimpulan tersebut ada beberapa poin penting sebagai berikut, klaster 1 dan klaster 4 merupakan klaster dengan anggota yang terdiri dari mahasiswa SAP angkatan 2014 dan 2015. Klaster - klaster ini memiliki performa pengerjaan skripsi yang standar dibandingkan klaster lainnya. Oleh karena itu klaster - klaster ini tidak perlu adanya perlakuan khusus. Selanjutnya adalah klaster 2, klaster ini memiliki jumlah anggota terbanyak, yaitu 184 mahasiswa dari total 372 data yang diteliti. Selain jumlah anggota yang banyak hal yang perlu diperhatikan pada klaster ini adalah performa pengerjaan skripsi nya. Klaster ini mengerjakan skripsi lebih lama durasi nya dibandingkan anggota – anggota klaster 3, yaitu klaster terbaik pada penelitian ini. Dan yang terakhir adalah klaster 3, jumlah anggota klaster ini hampir sama dengan klaster 2 yaitu berjumlah 173 mahasiswa yang semuanya angkatan 2015. Walaupun anggota klaster berjumlah banyak akan tetapi rata – rata durasi pengerjaan skripsi pada klaster ini merupakan yang tercepat. Oleh karena itu klaster 3 ini tidak perlu adanya perlakuan khusus dari pemegang keputusan program studi.

Klaster yang perlu diprioritaskan oleh pemegang keputusan program studi adalah klaster 2. Hal ini dikarenakan selain jumlah

anggota nya yang paling banyak dibandingkan dengan klaster lainnya, performa pengerjaan skripsi nya juga masih belum optimal. Pengerjaan skripsi yang optimal dapat dilihat di klaster 3, yang dimana kebetulan semua anggota nya angkatan 2015 / angkatan yang lebih muda. Jadi diharapkan kedepannya apabila terdapat mahasiswa yang memiliki kemiripan dari segi performa akademik nya ke klaster 2 maka mahasiswa tersebut dapat lebih diperhatikan lagi dalam pengerjaan skripsi nya dibandingkan dengan mahasiswa yang kemungkinan masuk ke klaster 3.

## 7. DAFTAR PUSTAKA

- Agarwal, S., Yadav, S., & Singh, K. (2012). K-means versus K-means ++ Clustering Technique, (March), 379–384.
- Ailon, N., Monteleoni, C., Jaiswal, R., & Monteleoni, C. (2009). Streaming k - means approximation. *Advances in Neural Information Processing Systems* 22, 10–18.
- Back, K. W. (2006). *Methods and Methods and Methods. Contemporary Psychology: A Journal of Reviews* (Vol. 21). <https://doi.org/10.1037/014836>
- Kaushik, M., & Mathur, B. (2014). Comparative Study of K-Means and Hierarchical Clustering Techniques. *International Journal of Software and Hardware Research in Engineering*, 2(6), 93–98.
- Nathan, A. J., & Scobell, A. (2012). Comparisons Between Data Clustering Algorithms. *The International Arab Journal of Information Technology*, 91(5), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Pechenizkiy, M., Trčka, N., Vasilyeva, E., Van Der Aalst, W., & De Bra, P. (2009). *Process Mining* online assessment data. *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, (January), 279–288.
- Premchaiswadi, W., & Porouhan, P. (2015). Process modeling and decision mining in a collaborative distance learning environment. *Decision Analytics*, 2(1), 0–34. <https://doi.org/10.1186/s40165-015-0015-5>

- Ryan Dwi Pambudi, Ahmad Afif Supianto, N. Y. S. (2019). Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan *Data Mining* Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2196, 3(3), 2194–2200.
- Syahrin, A. (2013). Implementasi algoritma k-means untuk klasterisasi mahasiswa berdasarkan prediksi waktu kelulusan skripsi. *UPN "Veteran" Jatim*, 1–23.
- Tadanai, O., & Tanuska, P. (2015). *Process Mining and practical usage. INES 2015 - IEEE 19th International Conference on Intelligent Engineering Systems, Proceedings*, 63–67. <https://doi.org/10.1109/INES.2015.7329758>
- Turban, E., Aronson, J. E., & Liang, T.-P. (2005). Decision Support Systems and Intelligent Systems (7th Ed). *Fenxi Huaxue*, 32(10). <https://doi.org/10.1017/CBO9781107415324.004>
- Vijayarani, S., Sakila, A., & Ramya, R. (2018). *Process Mining - A Comprehensive Review. International Journal of Computer Sciences and Engineering*, 6(7), 1108–1113. <https://doi.org/10.26438/ijcse/v6i7.11081113>