

Dikumpulkan di MS Teams bagian UAS 2022 max 16 Desember 2022 jam 22.00 WIB

1. Rancanglah Star Schema untuk Sistem Akademik di suatu universitas

a. Pemilihan Proses

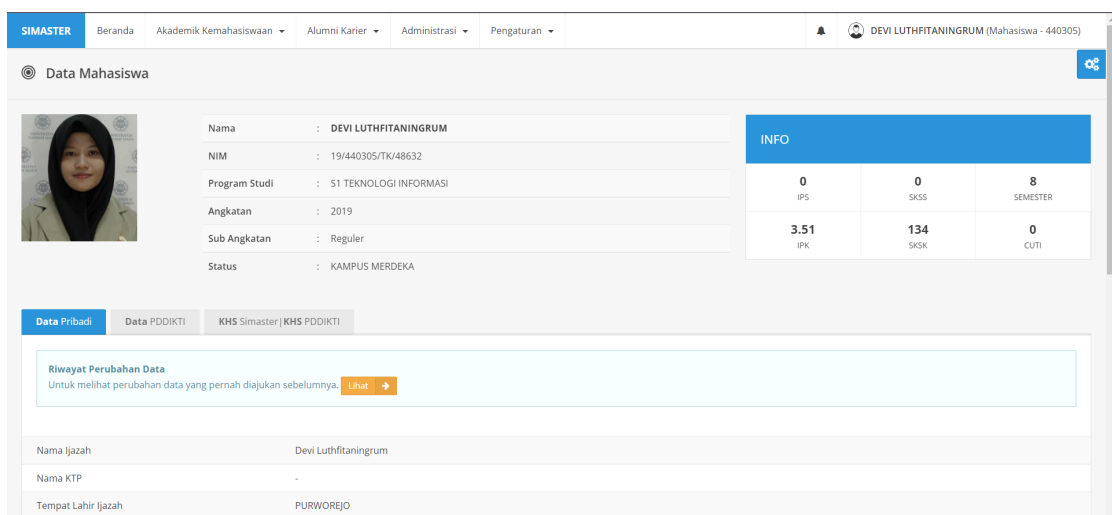
Data akademik merupakan data penting untuk merekam jejak aktivitas mahasiswa di universitas. Data perkuliahan dan data pengalaman yang pernah diambil di luar perkuliahan dapat digunakan untuk menganalisis keahlian yang dimiliki mahasiswa. Data ini dapat digunakan untuk melihat jejak *career path* dari mahasiswa tersebut.

Universitas meluluskan mahasiswa dengan jumlah meningkat setiap tahunnya. Lulusan universitas memiliki tanggung jawab sosial dan intelektual untuk menjadi anggota masyarakat yang memiliki kemampuan sehingga dapat menerapkan, mengembangkan dan menciptakan ilmu pengetahuan, teknologi, dan kesenian. Untuk mencapai hal tersebut, Dikti melakukan penyesuaian terhadap kompetensi lulusan universitas yaitu meliputi *hard skills* dan *soft skills*. Hal ini juga sebagai upaya menjawab permasalahan dominan yaitu kekhawatiran lulusan universitas menjadi pengangguran karena merasa belum dapat melihat potensi diri dan kurang memiliki pengalaman. Oleh karena itu, suatu universitas perlu melakukan upaya untuk membantu lulusannya dalam perencanaan karier yang baik.

b. Kebutuhan Data

Perancangan Star Schema untuk Sistem Akademik ini mengacu pada Aplikasi Simaster yang dimiliki Universitas Gadjah Mada. Aplikasi simaster dapat diakses melalui aplikasi web dan aplikasi *mobile*.

Berikut tampilan profil mahasiswa pada Aplikasi web Simaster.



SIMASTER Beranda Akademik Kemahasiswaan Alumni Karier Administrasi Pengaturan

DEVI LUTHFITANINGRUM (Mahasiswa - 440305)

Data Mahasiswa

INFO

| | | |
|------|------|----------|
| 0 | 0 | 8 |
| IPS | SKS | SEMESTER |
| 3.51 | 134 | 0 |
| IPK | SKSK | CUTI |

Data Pribadi Data PDDIKTI KHS Simaster | KHS PDDIKTI

Riwayat Perubahan Data
Untuk melihat perubahan data yang pernah diajukan sebelumnya. [Lihat](#)

| | |
|---------------------|----------------------|
| Nama Ijazah | Devi Luthfitaningrum |
| Nama KTP | - |
| Tempat Lahir Ijazah | PURWOREJO |

Dari data mahasiswa pada menu profil, dapat diidentifikasi bahwa kebutuhan datanya adalah Nama, NIM, Program Studi, Jenjang, dan Angkatan.

Sistem aplikasi simaster memiliki menu Akademik Kemahasiswaan dan Alumni Karier yang sangat mendukung mahasiswanya melakukan perencanaan karier. Berikut tampilan sub menu Hasil Studi yang terdapat pada menu Akademik Kemahasiswaan.

SIMASTER Beranda Akademik Kemahasiswaan Alumni Karier Administrasi Pengaturan

DEVI LUTHFITANINGRUM (Mahasiswa - 440305)

> Hasil Studi

Semester Semester Gasal 2022/2023

Cetak KHS

| No | Kode | Mata Kuliah | SKS | Ke | Nilai | Bobot | Nilai SKS | Proses Nilai |
|------------|-----------|--|-----|----|-------|-------|-----------|--------------|
| 1 | TIF215121 | Forensik Digital Kelas: A | 3 | 1 | | | | |
| 2 | TIF215132 | Big Data dan Analitik Kelas: A | 3 | 1 | | | | |
| 3 | TIF215213 | Pengalaman Pengguna Kelas: A | 3 | 1 | | | | |
| 4 | TIF214141 | Proyek Perancangan Teknologi Informasi 2 Kelas: A | 2 | 1 | | | | |
| 5 | TIF213241 | Proyek Senior Teknologi Informasi Kelas: A | 3 | 1 | | | | |
| Jumlah SKS | | | 14 | | | | 0.00 | |

Indeks Prestasi Sementara (IPS) = / =

20221

| | | | | | | | | |
|---|-----------|--|------|--|--|--|--|--|
| 1 | TIF213241 | Proyek Senior Teknologi Informasi | 3.00 | | | | | |
| 2 | TIF214141 | Proyek Perancangan Teknologi Informasi 2 | 2.00 | | | | | |
| 3 | TIF215121 | Forensik Digital | 3.00 | | | | | |
| 4 | TIF215132 | Big Data dan Analitik | 3.00 | | | | | |
| 5 | TIF215213 | Pengalaman Pengguna | 3.00 | | | | | |

Tentang
SIMASTER UGM

UNIVERSITAS GADJAH MADA
Bulaksumur Yogyakarta 55281
E-mail: dssdi@ugm.ac.id

Copyright © Universitas Gadjah Mada. All rights reserved.

Berdasarkan aplikasi simaster, dapat diidentifikasi bahwa untuk merekam hasil studi dibutuhkan data periode semester dan tahun, kode mata kuliah, mata kuliah, SKS, dan nilai.

Berikut tampilan sub menu Curriculum Vitae yang merekam jejak pengalaman mahasiswa pada menu Alumni Karier.

SIMASTER Beranda Akademik Kemahasiswaan Alumni Karier Administrasi Pengaturan

DEVI LUTHFITANINGRUM (Mahasiswa - 440305)

Curriculum Vitae

Curriculum Vitae

Pendidikan Pekerjaan Organisasi Penghargaan Pelatihan Beasiswa **Keahlian** Bahasa Wirausaha

+ Tambah Keahlian

| No. | Nama Keahlian | Tingkat | Publish | Aksi |
|----------------------------|---------------|---------|---------|------|
| -- Data tidak ditemukan -- | | | | |

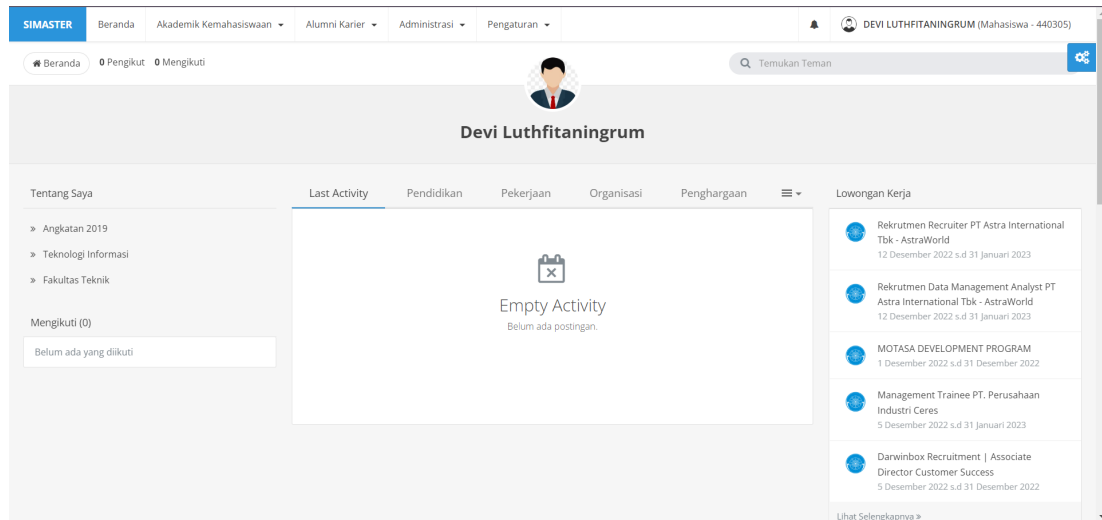
Tentang
SIMASTER UGM

UNIVERSITAS GADJAH MADA
Bulaksumur Yogyakarta 55281
E-mail: dssdi@ugm.ac.id

Copyright © Universitas Gadjah Mada. All rights reserved.

Data yang dicantumkan pada sistem mengenai pengalaman berupa nomor id, nama pengalaman seperti nama pelatihan yang diikuti, pengalaman posisi magang, serta deskripsi penjelasan dan periode waktu dari pengalaman tersebut.

Berikut tampilan sub menu baru di aplikasi simaster yaitu simfoni, terletak pada menu Alumni Karier. Sub menu simfoni berisi data tentang mahasiswa terkait nama, angkatan, prodi, fakultas, dan aktivitas yang diinput di sub menu Curriculum Vitae.



Proses yang dipilih dalam perancangan star schema yaitu proses perencanaan karier. Perancangan yang dilakukan akan mewakili sub menu Simfoni, yaitu perekaman aktivitas mahasiswa. Meskipun sedikit berbeda karena adanya data tambahan yang dibutuhkan seperti data hasil studi.

c. Pemilihan Grain

Memilih *grain* berarti menentukan secara tepat apa yang diwakili pada tabel fakta. Sesuai dengan pemilihan proses sebelumnya, *grain* dari tabel fakta adalah aktivitas mahasiswa di dalam ataupun di luar perkuliahan.

d. Identifikasi dan penyesuaian dimensi

| | | |
|---------------|---------------|--|
| Mahasiswa | dim_mahasiswa | ID mahasiswa, nama mahasiswa, nim, angkatan |
| Program Studi | dim_prodi | ID prodi, prodi, departemen, fakultas, jenjang |
| Periode | dim_periode | ID periode, periode, tahun |
| Mata kuliah | dim_matkul | ID matkul, nama matkul, SKS, nilai |

| | | |
|------------|----------------|--|
| Pengalaman | dim_pengalaman | ID pengalaman, tipe pengalaman, judul pengalaman, deskripsi, durasi |
|------------|----------------|--|

e. Pemilihan fakta

Disesuaikan dengan *grain* dan dimensi, fakta yang mewakili aktivitas mahasiswa di dalam ataupun di luar perkuliahan yaitu fact_aktivitas. Fact_aktivitas menyimpan data mahasiswa seperti IPK, sks yang diambil, dan keahlian.

| | | |
|-----------|----------------|--|
| Aktivitas | fact_aktivitas | ID aktivitas, ID mahasiswa, ID prodi, ID periode, ID matkul, ID pengalaman, IPK, sks_ambil, keahlian |
|-----------|----------------|--|

f. Menyimpan perhitungan awal dalam tabel fakta

Dalam perekaman aktivitas mahasiswa tidak ada penyimpanan khusus dalam basis data, namun data akan didapatkan melalui *query* dan *data mining* atau *text mining* yang dijalankan nantinya. Operasi *query* dapat dijalankan untuk menghitung IPK atau nilai IP kumulatif berdasarkan data nilai dan bobot SKS setiap mata kuliah serta jumlah total sks yang sudah diambil mahasiswa. *Data mining* dan *text mining* yang dilakukan berupa *classification* dan *clustering* yang menghasilkan pengelompokan mahasiswa ke dalam beberapa kategori keahlian.

g. Melihat kembali tabel dimensi

dim_mahasiswa

| | | | |
|----------------|----------------|---------|--------------------------------|
| ID mahasiswa | id_mahasiswa | int | Nomor Induk Universitas (unik) |
| nama mahasiswa | nama_mahasiswa | varchar | Nama lengkap mahasiswa |
| nim | nim | varchar | Nomor Induk Mahasiswa lengkap |
| angkatan | angkatan | int | Tahun masuk mahasiswa |

dim_prodi

| | | | |
|------------|------------|---------|-------------------------|
| ID prodi | id_prodi | varchar | Kode unik program studi |
| prodi | prodi | varchar | Nama program studi |
| departemen | departemen | varchar | Nama departemen |
| fakultas | fakultas | varchar | Nama fakultas |
| jenjang | jenjang | varchar | Jenjang pendidikan |

dim_periode

| | | | |
|------------|------------|---------|--------------------------|
| ID periode | id_periode | int | Nomor unik periode |
| periode | periode | varchar | periode gasal atau genap |
| tahun | tahun | int | tahun |

dim_matkul

| | | | |
|-------------|--------------|---------|------------------------|
| ID matkul | id_matkul | varchar | kode unik mata kuliah |
| nama matkul | matkul | varchar | Nama mata kuliah |
| SKS | sks_matkul | float | jumlah sks mata kuliah |
| nilai | nilai_matkul | float | nilai mahasiswa |

dim_pengalaman

| | | | |
|----------------------|------------------|---------|--|
| ID pengalaman | id_pengalaman | varchar | kode unik pengalaman |
| tipe pengalaman | tipe_pengalaman | varchar | Magang/PartTime/Pelatihan/Riset/ Proyek/Skripsi |
| judul pengalaman | judul_pengalaman | varchar | jabatan atau judul kegiatan/riset/proyek/skripsi |
| deskripsi pengalaman | desk_pengalaman | varchar | deskripsi pengalaman yang dilakukan dan keahlian yg dimiliki |
| durasi | durasi | varchar | durasi pengalaman |

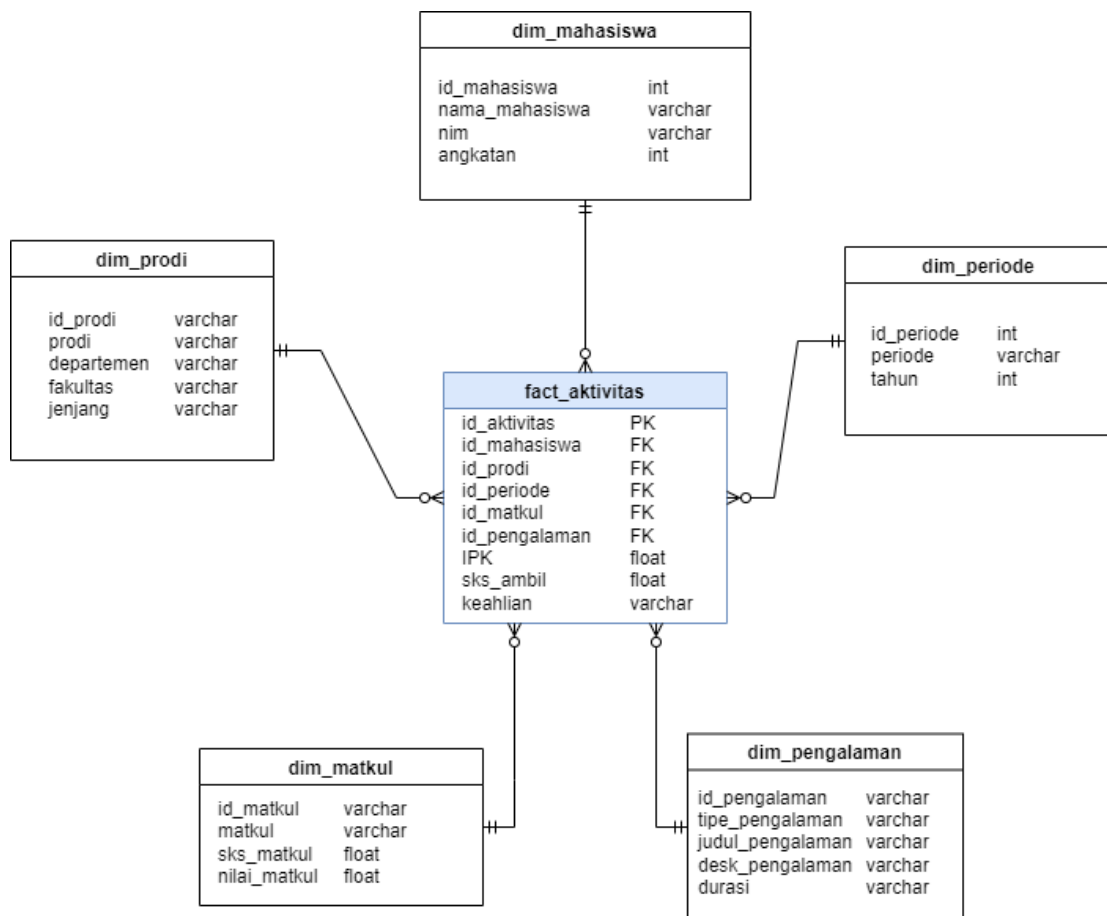
h. Memilih durasi database

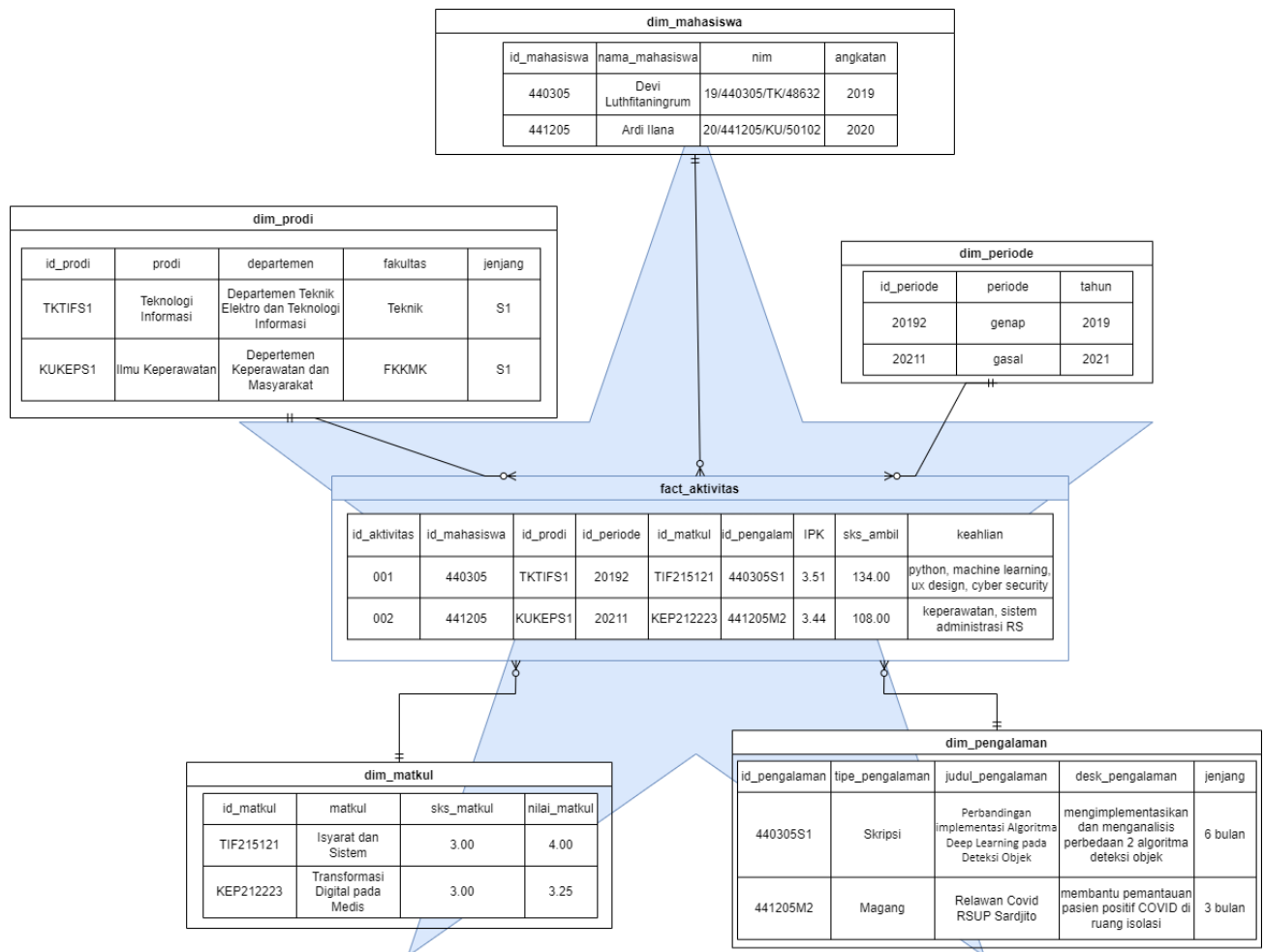
Durasi database yang ditetapkan adalah 5 tahun sehingga sivitas akademik bisa melihat aktivitas mahasiswa selama 5 tahun. Hal ini mempertimbangkan rata-rata periode kelulusan mahasiswa adalah 4 hingga 5 tahun. Jadi sivitas akademik dapat menganalisis data dengan lebih mudah karena data keseluruhan selama masa studi masih tersimpan.

i. Menelusuri perubahan dimensi

- Dimensi yang berubah ditimpa ulang
- Atribut dari dimensi yang berubah mengakibatkan munculnya *record baru* (tidak ada fitur *update*)
- Atribut dimensi mengakibatkan atribut lain muncul menjadi nilai lama dan baru

j. Perancangan model data Star Schema





2. Jika kita akan melakukan data mining **classification** atau **clustering** data pada soal nomor 1, maka jelaskan data mana yang dipakai, bagaimana prosesnya, dan outputnya

Analisis data mining meliputi tahapan proses *Knowledge Discovery in Database (KDD)*. KDD adalah proses menemukan informasi berguna dan pola di data. Berikut langkah-langkah dalam proses KDD:

1. memahami *prior knowledge* dan tujuan analisis
2. membuat dataset target: *data selection*
3. *Pre-processing* dan *cleaning data*
4. *Transformation*
5. **Data Mining**
6. Evaluasi pola dan presentasi *discovered knowledge*

Classification: memetakan data ke grup atau kelas yang sudah ditentukan

- *Supervised learning*
- *Pattern recognition*
- *Predictive Modeling*

Classification data pada soal nomor 1 dapat dilakukan dalam:

Klasifikasi untuk prediksi predikat IPK kelulusan mahasiswa

Adapun kategori Predikat IPK kelulusan universitas adalah sebagai berikut:

| Range IP | Predikat |
|-----------|------------------|
| 2.50-2.74 | Memuaskan |
| 2.75-3.50 | Sangat memuaskan |
| 3.51-3.79 | Cum laude |
| 3.80-3.99 | Magna cum laude |
| 4.00 | Summa cum laude |

Data selection, pre-processing, cleaning data, dan transformasi data

Data bisa didapatkan dengan query data warehouse dan contoh output perhitungan sebagai berikut:

```
SELECT id_mahasiswa as "NIU", nama_mahasiswa as "Nama", nilai_matkul as "nilai",  
sks_matkul as "sks", nilai*sks as "total_nilai"  
FROM fact_aktivitas, dim_mahasiswa, dim_matkul  
GROUP BY(NIU)
```

| NIU | Nama | nilai | sks | total_nilai |
|--------|-------------|-------|-------|-------------|
| 440305 | Devi Luthfi | 3.25 | 2.00 | 6.50 |
| 440305 | Devi Luthfi | 3.00 | 3.00 | 9.00 |
| 440305 | Devi Luthfi | 3.00 | 4.00 | 12.00 |
| 440305 | Devi Luthfi | 4.00 | 3.00 | 12.00 |
| 440305 | Devi Luthfi | 3.25 | 2.00 | 6.50 |
| null | null | null | 14.00 | 46.00 |

$IPS/IPK = 46.00/14.00 = 3.3$ (sangat memuaskan)

Berdasarkan tahap *data selection, pre-processing, cleaning data, dan transformasi data* didapatkan konsep dataset yang dibutuhkan untuk klasifikasi mahasiswa berdasarkan kelompok predikat IPK kelulusannya adalah seperti tabel berikut:

| NIU | Nama | IPS 1 | IPS 2 | IPS 3 | IPS 4 | IPS 5 | IPS 6 | IPS 7 | IPS 8 | IPK | Predikat |
|--------|-------------|----------|----------|----------|----------|----------|----------|----------|----------|-----|------------------|
| 440305 | Devi Luthfi | 3.3 | 3.7 | 3.4 | 3.5 | 3.3 | 3.6 | 3.8 | 4.0 | 3.6 | cumlaude |
| 440402 | Nadya Rosa | 2.2 | 1.5 | 2.2 | 1.9 | 3.0 | 2.4 | 2.6 | 3.2 | 2.5 | memuaskan |
| 440403 | Fazerian | 3.0 | 3.6 | 2.2 | 2.3 | 2.5 | 2.8 | 3.0 | 2.9 | 2.8 | sangat memuaskan |
| 440450 | David Yuli | 3.8 | 3.4 | 3.9 | 3.8 | 3.8 | 3.9 | 3.8 | 3.8 | 3.8 | magna cum laude |
| 440451 | Mufid Hani | 2.0 | 1.5 | 1.8 | 2 | 2.2 | 2.3 | 2.4 | 3.2 | 2.6 | memuaskan |

Dataset berisi nim, nama, dan IP setiap semester di-*split* menjadi dataset *training* dan *test*. Algoritma yang biasa digunakan untuk klasifikasi adalah algoritma *K-Nearest Neighbor*. Pada algoritma ini, klasifikasi yang sudah ada di data *training* dapat digunakan untuk prediksi data yang belum terklasifikasi. Algoritma ini menemukan kelompok K dalam *data training* yang paling dekat dengan *data test* dan dilanjutkan pencarian label kelas tertentu. Pengujian nilai k terbaik berdasarkan akurasi tertinggi dilakukan menggunakan data *test*. Hasil klasifikasi dengan algoritma dibandingkan dengan data yang sudah diklasifikasi. Berikut contoh proses pencarian nilai k dengan akurasi tertinggi:

| Nilai k | Nilai benar | Nilai salah | Persentase akurasi |
|---------|-------------|-------------|--------------------|
| 1 | 198 | 52 | 78,6% |
| 3 | 202 | 50 | 80,1% |
| 5 | 228 | 22 | 91,3% |

Maka algoritma K-Nearest Neighbor yang terbaik menggunakan nilai K=5.

Prediksi klasifikasi dilakukan ketika mahasiswa belum lulus (nilai IPS belum terpenuhi 8 semester) untuk *forecast* predikat IPK kelulusan. Dengan mengetahui prediksi predikat IPK lebih awal, hal ini akan membantu mahasiswa dalam menyusun strategi pencapaian target IPK yang diinginkan. Sehingga mahasiswa akan termotivasi untuk meningkatkan nilai di semester berikutnya.

Clustering: memetakan data yang memiliki kemiripan ke grup atau kelas yang belum diketahui

- *Unsupervised learning*
- *Segmentation*
- *Partitioning*

Clustering data pada soal nomor 1 dapat dilakukan dalam:

Clustering mata kuliah peminatan

Agar lebih mudah dipahami, dalam proses kali ini akan dilakukan *clustering* pada mata kuliah peminatan Teknologi Informasi. Program studi Teknologi Informasi memiliki 12 mata kuliah peminatan sebagai berikut:

Software Engineer

- TIF215111: Pengembangan Aplikasi Permainan
- TIF215112: Pengembangan Aplikasi Piranti Bergerak
- TIF215113: Arsitektur Perangkat Lunak
- TIF215114: Pengujian Perangkat Lunak

Data Engineer

- TIF215131: Pemrosesan Bahasa Alami
- TIF215132: Big Data dan Analitik
- TIF215133: Sistem Pendukung Keputusan
- TIF215134: Teknik Kompresi Data

Network & Security Engineer

- TIF215121: Forensik Digital
- TIF215122: Peretasan Beretika
- TIF215123: Keamanan dan Integritas Data
- TIF215124: Sistem Komunikasi Bergerak

Data selection, pre-processing, cleaning data, dan transformasi data

Data bisa didapatkan dengan query data warehouse dan contoh output perhitungan sebagai berikut:

```
SELECT id_matkul, COUNT(id_mahasiswa) as "jumlah_peserta"  
FROM fact_aktivitas  
GROUP BY (id_matkul)
```

Berdasarkan tahap *data selection, pre-processing, cleaning data, dan transformasi data* didapatkan konsep dataset yang dibutuhkan untuk *clustering* mata kuliah peminatan adalah seperti tabel berikut:

| id_matkul | jumlah_peserta |
|-----------|----------------|
| TIF215111 | 33 |
| TIF215112 | 12 |
| TIF215113 | 80 |
| TIF215114 | 30 |
| TIF215121 | 30 |
| TIF215122 | 15 |
| TIF215123 | 60 |

| | |
|-----------|----|
| TIF215124 | 12 |
| TIF215131 | 30 |
| TIF215132 | 45 |
| TIF215133 | 30 |
| TIF215134 | 12 |

Algoritma yang biasa digunakan untuk *clustering* adalah algoritma *K-Means*. Algoritma ini diawali dengan menentukan banyaknya *cluster* yang diinginkan dari dataset yang ada. Pada percobaan kali ini, ditentukan nilai $k=3$ artinya akan terbentuk 3 kluster yaitu C1, C2, dan C3. Pengelompokan berdasarkan variabel mata kuliah peminatan kurang diminati, cukup diminati, dan banyak diminati. Selanjutnya ditentukan centroid awal secara random.

| Centroid | jumlah_peserta |
|----------|----------------|
| C1 | 12 |
| C2 | 30 |
| C3 | 80 |

Lalu dengan rumus Euclidean Distance dilakukan perhitungan jarak data dengan *centroid*.

| id_matkul | jumlah_peserta | jarak1 | jarak2 | jarak3 | cluster |
|-----------|----------------|--------|--------|--------|---------|
| TIF215111 | 33 | 21 | 3 | 47 | C2 |
| TIF215112 | 12 | 0 | 18 | 68 | C1 |
| TIF215113 | 80 | 68 | 50 | 0 | C3 |
| TIF215114 | 30 | 18 | 0 | 50 | C2 |
| TIF215121 | 30 | 18 | 0 | 50 | C2 |
| TIF215122 | 15 | 3 | 15 | 65 | C1 |
| TIF215123 | 60 | 48 | 30 | 20 | C3 |
| TIF215124 | 12 | 0 | 18 | 68 | C1 |
| TIF215131 | 30 | 18 | 0 | 50 | C2 |
| TIF215132 | 45 | 33 | 15 | 35 | C2 |
| TIF215133 | 30 | 18 | 0 | 50 | C2 |

| | | | | | |
|-----------|----|---|----|----|----|
| TIF215134 | 12 | 0 | 18 | 68 | C1 |
|-----------|----|---|----|----|----|

Klaster didapatkan dari centroid yang memiliki jarak minimum dengan data jumlah peserta. Algoritma K-Means dilakukan secara iterasi dengan perubahan centroid hingga anggota klaster tidak berubah. Nilai centroid pada iterasi berikutnya dihitung dari rata-rata data dari anggota klaster pada iterasi sebelumnya.

Contoh centroid $C1 = (12+15+12+12)/4 = 13$, $C2 = 198/6 = 33$, $C3 = (80+60)/2 = 70$

Dilanjutkan perhitungan jarak data dengan centroid seperti sebelumnya.

Dari iterasi 1, didapatkan klaster seperti berikut:

Mata kuliah kurang diminati: TIF215112, TIF215122, TIF215124, TIF215134

Mata kuliah cukup diminati: TIF215111, TIF215114, TIF215121, TIF215131, TIF215132, TIF215133

Mata kuliah banyak diminati: TIF215113, TIF215123

3. Jika kita akan melakukan **forecast** data pada soal nomor 1, maka jelaskan data mana yang dipakai, bagaimana prosesnya, dan *outputnya*

Forecast IPS (Indeks Prestasi Semester) untuk tiap semester selanjutnya perlu dilakukan agar menjadi motivasi bagi mahasiswa dalam mencapai target IPK. Karena nilai IPK tidak bisa ditingkatkan secara instan, melainkan perlu strategi belajar dari awal masuk agar IPS diharapkan stabil dan memuaskan. IPS merupakan variabel respon terhadap waktu yaitu tiap periode semester sehingga dapat di-*forecast* dengan time series forecast. Dipertimbangkan dari nilai IPS yang seharusnya tidak membentuk suatu *trend* melainkan terjadi perubahan secara *smooth*, maka metode yang tepat adalah *smoothing methods*. Salah satu metode *smoothing* yang tepat untuk *forecast* nilai IPS adalah metode *exponential smoothing*. Pada *exponential smoothing*, nilai yang digunakan untuk *forecast* IPS selanjutnya adalah nilai IPS dan nilai *forecast* lalu. Hal ini dapat meminimalisasi faktor-faktor lain yang mempengaruhi IPS di periode tertentu (terbentuk *outlier*) pada perhitungan *forecast*. Nilai IPS periode lampau diberi bobot ω dimana $0 \leq \omega \leq 1$.

$\text{Forecast}(t) = \text{Forecast}(t-1) + \omega * \{\text{Actual}(t-1) - \text{Forecast}(t-1)\}$

Data yang dipakai untuk *forecast* merupakan data IPS yang bisa didapat dari query seperti pada soal nomor 2 yang diambil di setiap periode semester.

```
SELECT id_mahasiswa as "NIU", nilai_matkul as "nilai", sks_matkul as "sks", nilai*sks  
as "total_nilai
```

```
FROM fact_aktivitas, dim_matkul
```

```
WHERE id_mahasiswa=440305 AND id_periode=20191
```

```
GROUP BY(NIU)
```

| NIU | id_periode | nilai | sks | total_nilai |
|--------|------------|-------|------|-------------|
| 440305 | 20191 | 3.25 | 2.00 | 6.50 |
| 440305 | 20191 | 3.00 | 3.00 | 9.00 |
| 440305 | 20191 | 3.00 | 4.00 | 12.00 |

| | | | | |
|--------|-------|------|-------|-------|
| 440305 | 20191 | 4.00 | 3.00 | 12.00 |
| 440305 | 20191 | 3.25 | 2.00 | 6.50 |
| null | null | null | 14.00 | 46.00 |

$$\text{IPS 1} = 46.00/14.00 = 3.33$$

Query tersebut dilakukan setiap periode semester dan dilakukan ke tiap data mahasiswa

Berikut contoh perhitungan *forecast* dengan bobot $\omega=0.2$:

| Semester | Data Real | Forecast (Exponential smoothing) $\omega=0.2$ |
|----------|-----------|---|
| 1 | 3.33 | 3.33 (disamakan data awal) |
| 2 | 3.56 | $3.33+0.2*(3.33-3.33) = 3.33$ |
| 3 | 3.39 | $3.33+0.2*(3.56-3.33) = 3.38$ |
| 4 | 3.48 | $3.38+0.2*(3.39-3.38) = 3.38$ |
| 5 | 3.34 | $3.38+0.2*(3.48-3.38) = 3.4$ |
| 6 | 3.64 | $3.4+0.2*(3.34-3.4) = 3.41$ |
| 7 | - | $3.41+0.2*(3.64-3.41) = 3.46$ |

Berdasarkan *forecast exponential smoothing*, didapatkan kemungkinan IP Semester 7 adalah 3.46. Hasil *forecast* dari semester-semester sebelumnya tidak terlalu jauh dibanding data asli sehingga dapat dikatakan *forecast* cukup mewakili data IP semester mahasiswa dengan NIU 440305.

4. Apakah data pada soal nomor 1 memungkinkan untuk dilakukan association rule mining

Association rule menemukan hubungan antar-data. Hal ini bisa dilakukan untuk menganalisis hubungan antar mata kuliah peminatan yaitu kemungkinan dimana mahasiswa yang mengambil mata kuliah peminatan tertentu juga akan mengambil mata kuliah peminatan yang lain. Data yang dibutuhkan untuk operasi ini adalah data mahasiswa dan data mata kuliah yang diambil.

SELECT id_mahasiswa as "NIU", id_matkul

FROM fact_aktivitas

GROUP BY (id_mahasiswa)

| NIU | id_matkul |
|--------|-----------|
| 440303 | TIF215131 |

| | |
|--------|-----------|
| 440303 | TIF215132 |
| 440303 | TIF215133 |
| 440305 | TIF215123 |
| 440305 | TIF215121 |
| 440305 | TIF215132 |
| 440307 | TIF215123 |
| 440307 | TIF215131 |
| 440307 | TIF215132 |
| 440309 | TIF215111 |
| 440309 | TIF215112 |
| 440309 | TIF215121 |

Untuk melakukan association rule, data diatas perlu diubah ke dalam bentuk cross-tabulated seperti berikut:

| Matkul | TIF215111 | TIF215112 | TIF215121 | TIF215123 | TIF215131 | TIF215132 | TIF215133 |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| TIF215111 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| TIF215112 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| TIF215121 | 1 | 1 | 2 | 1 | 0 | 0 | 0 |
| TIF215123 | 0 | 0 | 1 | 2 | 1 | 1 | 0 |
| TIF215131 | 0 | 0 | 0 | 1 | 2 | 1 | 1 |
| TIF215132 | 0 | 0 | 0 | 2 | 2 | 3 | 1 |
| TIF215133 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Mata kuliah TIF215131 dan TIF215132 banyak diambil bersamaan oleh mahasiswa yang sama, sedangkan mata kuliah yang hampir diambil semua mahasiswa di atas adalah mata kuliah TIF215132.

5. Jika kita akan melakukan text mining yang berhubungan dengan sistem akademik di suatu universitas, data apa yang memungkinkan dipakai, jelaskan bagaimana prosesnya, dan *outputnya*

Data yang dapat dipakai untuk *text mining* yaitu data pada *dim_pengalaman*. Data yang dapat digunakan untuk *text mining* adalah *judul_pengalaman* dan *desk_pengalaman*. Text mining dapat dilakukan untuk mencari *skill* atau keahlian yang mahasiswa miliki berdasarkan pengalaman seperti magang, proyek, riset, dan lain-lain. Hal ini sangat dibutuhkan untuk mahasiswa maupun industri saat ini, dimana mahasiswa perlu mempublikasikan keahliannya agar dapat direkrut oleh perusahaan industri atau lapangan kerja dan proyek lainnya. Dengan adanya sistem akademik dengan text mining dan perancangan data model sebelumnya, diharapkan universitas dapat membantu mahasiswa menemukan keahlian dan potensi dirinya.

Text mining kali ini dilakukan dengan metode pencarian kata inti. Hal ini memerlukan tahap pembobotan dokumen untuk menemukan kata inti dari kalimat. Sumber data pada *text mining* adalah kumpulan teks tidak atau semi struktur. Berikut tahapan *text processing* yang bertujuan mempersiapkan teks menjadi data terstruktur dan dapat diproses.

1. **Cleansing**, proses memperbaiki atau menghapus data yang rusak dan tidak lengkap. Karena teks yang digunakan adalah data input di sistem akademik, diharapkan karakteristik kalimat cukup baku dan semi struktur sehingga tidak banyak *noise* dan kata tidak baku seperti di *twitter*.
2. **Case Folding**, proses mengubah huruf 'a' sampai dengan 'z' dalam teks menjadi huruf kecil.
3. **Tokenizing**, memotong string input berdasarkan tiap kata yang menyusunnya.
4. **Filtering**, mengambil kata-kata penting dari hasil tokenizing menggunakan algoritma stopword removal dengan membuang kata yang kurang penting.
5. **Stemming**, mencari *root* kata dari tiap kata hasil filtering.

Tahap selanjutnya adalah metode *Term Frequency-Invers Document Frequency* (TF-IDF). Metode ini adalah cara pemberian bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan 2 konsep penghitungan bobot yaitu frekuensi kemunculan kata pada kalimat dan frekuensi banyaknya dokumen dimana suatu kata muncul. Bobot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu teks dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen. Hal ini tepat untuk menemukan potensi diri dan keahlian mahasiswa yang unik dari pengalaman mahasiswa. Karena kata istilah keahlian yang sering muncul di judul atau deskripsi pengalaman mahasiswa akan menjadi keahlian mahasiswa tersebut. Selain itu, istilah keahlian yang banyak dimiliki oleh orang lain akan memiliki bobot yang kecil. Jadi, bobot setiap kata adalah:

$$\omega = \text{TF} * \text{IDF}$$

Contoh text mining:

| | | |
|--------|---|--|
| 440305 | UNESCO UNITWIN Online Program Training | I lead the team to design Software Interface based natural language processing. In the last session, I am present our report and I can be Best Presenter of Innovation Idea Presentation |
| 440305 | Staff of Research and Technology | I lead the software requirement analysis and design team to designing technology device |

| | | |
|--|----------|---|
| | Division | and network architecture then present it to other division. I am communicate and contact technology stakeholders. |
|--|----------|---|

Berikut contoh perhitungan bobot TF-IDF

| Kata kunci keahlian | Bobot |
|---------------------|-----------------|
| Lead | $2 * 0,2 = 0,4$ |
| Design | $3 * 0,2 = 0,6$ |
| Software | $2 * 0,2 = 0,4$ |
| Technology | $2 * 0,5 = 1$ |
| Present | $3 * 0,2 = 0,6$ |

Kata kunci keahlian dengan bobot tertinggi dari mahasiswa NIU 440305 adalah **technology** karena kata tersebut sering muncul dalam teks kalimat tetapi tidak banyak muncul di dokumen lain. Jadi, dapat disimpulkan bahwa **career path** mahasiswa tersebut berkaitan dengan **technology**. Pencarian kata kunci keahlian perlu dilakukan lagi untuk menemukan keahlian lain yang dimiliki oleh mahasiswa, minimal 5 kata kunci keahlian yang biasanya dapat mewakili untuk digunakan dalam keperluan lain seperti melamar beasiswa, pekerjaan, magang, dan lain-lain. Untuk melengkapi *text-mining* maka perlu dilanjutkan K-Means *clustering* seperti pada soal nomor 2.

Kesimpulan

Penyimpanan dan penambahan data melalui query OLAP dan berbagai teknik *Knowledge Discovery in Database* sangat bermanfaat dalam suatu organisasi untuk mengidentifikasi berbagai hal yang dibutuhkan. Banyaknya data tentu menjadi tantangan dalam menganalisis keseluruhan data. Namun dengan banyaknya data yang dimiliki juga tersimpan banyak manfaat dalam upaya **forecast** kebutuhan pengguna sistem dari berbagai aspek.

Perancangan Data pada dokumen ini belum diaplikasikan secara keseluruhan pada Aplikasi Simaster, hal ini mungkin dapat dikembangkan untuk sistem ke depannya agar dapat membantu personalisasi keahlian bagi setiap mahasiswa dalam persiapan karirnya. Perancangan data dan pembahasan teknik KDD pada dokumen ini juga masih dapat dikembangkan lagi untuk berbagai keperluan yang dibutuhkan dalam sistem akademik universitas terutama Simaster UGM.

Referensi

- Budiman, I., Muliadi, & Ramadina, R. (2015, April). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. *Jurnal Jupiter*, 7(1). <https://media.neliti.com/media/publications/289138-penerapan-fungsi-data-mining-klasifikasi-c107fb0c.pdf>
- Fadilah, U., Winarno, W. W., & Amborowati, A. (2016, Juli). Perancangan Data Warehouse Untuk Sistem Akademik STMIK Kadiri. *Jurnal Ilmiah SISFOTENIKA*, 6(2), 217-228. <https://sisfotenika.stmikpontianak.ac.id/index.php/ST/article/view/119/119>
- Implementasi Star Schema dalam Perancangan Data Warehouse Akademik Perguruan Tinggi. (2021, Oktober). *CAKRAWALA - Repositori IMWI*, 4(2). <https://cakrawala.imwi.ac.id/index.php/cakrawala/article/view/91/70>
- Indriyani, F., & Irfiani, E. (2019, November). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K-Means Method). *JUITA: Jurnal Informatika*, 7(2), 109-113. https://www.researchgate.net/publication/337599172_Clustering_Data_Penjualan_pada_Toko_Perlengkapan_Outdoor_Menggunakan_Metode_K-Means
- IRWANSYAH, E. (2017, March 9). *CLUSTERING*. School of Computer Science | BINUS University. Retrieved December 16, 2022, from <https://socs.binus.ac.id/2017/03/09/clustering/>
- Kambey, G. E.I., Sengkey, R., & Jacobus, A. (2020, April-Juni). Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia. *Jurnal Teknik Informatika*, 15(2), 75-82.
- Noviansyah, M. R., Rismawan, T., & Midyanti, D. M. (2018). PENERAPAN DATA MINING MENGGUNAKAN METODE K-NEAREST NEIGHBOR UNTUK KLASIFIKASI INDEKS CUACA KEBAKARAN BERDASARKAN DATA AWS (AUTOMATIC WEATHER STATION) (STUDI KASUS: KABUPATEN KUBU RAYA). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48-56. <https://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672/75676577366>
- Nurafifah, H. (2017). KLASIFIKASI MINAT STUDI MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYESIAN CLASSIFIERS (NBC). *Simki-Techsain*, 01(06), 1-7. http://simki.unpkediri.ac.id/mahasiswa/file_artikel/2017/ffb9db9392a901f249629594efda066d.pdf
- Ozora, D., Suharti, L., & Sirine, H. (2016). *POTRET PERENCANAAN KARIR PADA MAHASISWA (Studi terhadap Mahasiswa di Sebuah Perguruan Tinggi di Jawa Tengah)*. media.neliti.com.

<https://media.neliti.com/media/publications/171974-ID-potret-perencanaan-karir-pada-mahasiswa.pdf>

- Siregar, M. H. (2018, Desember). KLASTERISASI PENJUALAN ALAT-ALAT BANGUNAN MENGGUNAKAN METODE K-MEANS (STUDI KASUS DI TOKO ADI BANGUNAN). *JURNAL TEKNOLOGI DAN OPEN SOURCE*, 1(2), 83-91. <https://media.neliti.com/media/publications/284729-data-mining-klasterisasi-penjualan-alat-b873003e.pdf>
- Suaibi, R., Lusiana, D., & Daryanto. (n.d.). APLIKASI TEKS MINING UNTUK AUTOMASI PENCARIAN KALIMAT INTI DALAM DOKUMEN TUNGGAL BERBAHASA INDONESIA DENGAN METODE Term Frequency-Inverse Document Frequency (TF-IDF). *repository.unmuhjember*. <http://repository.unmuhjember.ac.id/2734/9/JURNAL.pdf>
- Ulfah, R. F. (2017). 1 ANALISIS PERENCANAAN KARIER BERDASARKAN KOMPETENSI MAHASISWA PADA MAHASISWA PROGRAM STUDI PENDIDIKAN AKUNTANSI FAKULTAS KEGURU. Retrieved December 16, 2022, from <http://eprints.ums.ac.id/52865/2/NASKAH%20PUBLIKASI.pdf>