

# EXPLORATORY DATA ANALYSIS PADA DATABASE SISTEM INFORMASI AKADEMIK STIKOM DINAMIKA BANGSA

Beny

Teknik Informatika, STIKOM Dinamika Bangsa, Jambi

Jl. Jendral Sudirman Thehok, 0741-35095

E-mail: beny@stikom-db.ac.id

## ABSTRAK

STIKOM Dinamika Bangsa memiliki pangkalan data yang menyimpan sebagian besar transaksi data akademik di tiap semester. Data-data yang cukup besar ini mengandung informasi yang siap digali. Untuk salah satu informasi yang dapat diambil dari pangkalan data tersebut adalah pola-pola dan relasi antar variabel. Untuk mendapatkan informasi itu bisa melalui proses eksplorasi data. Pada penelitian ini penulis melakukan proses exploratory data analysis pada pangkalan data sistem informasi Sekolah Tinggi Ilmu Kompter Dinamia Bangsa untuk mendapatkan gambaran pola-pola, sebaran data, dan korelasi antar variabel. Hasil dari penelitian ini terbentuklah data yang lebih rapih, tergambarannya sebaran data nilai ipk berdasarkan beberapa variabel lainnya, dan teridentifikasinya beberapa variabel yang memiliki korelasi. Dari hasil yang didapatkan tersebut dapat disimpulkan bahwa variabel angkatan terhadap rata-rata ipk menunjukkan nilai korelasi Pearson Product Moment sebesar 0.91 dengan Confidence Interval 95% dan p-value 0.0003, ini menandakan korelasi yang secara statistic signifikan.

*Kata kunci: data besar, exploratory data analysis, data akademik*

## ABSTRACT

STIKOM Dinamika Bangsa has database for storing its main academic transaction data for each semester. The data are big enough to contain information which is ready to be mined. One of the information that can be extracted from the database is patterns and correlation between variabel. To get these information, we can do exploratory data analysis. In this research, we did exploratory data analysis process to get patterns, data spread, and correlation between variables. The result from this research is we managed to get tidy data, visualization of academic score data to other variables, and we identified some variables which has correlation to each other. We found that year class correlates with gpa, with 0.91 Pearson Product Moment, 95% Confidence Interfal, and 0.0003 p-value, this means the correlation is statistically significant.

*Keywords: big data, exploratory data analysis, academic data*

## 1. PENDAHULUAN

Pada era big data saat ini terdapat banyak informasi yang bisa digali dari berbagai bidang [1], salah satunya adalah bidang pendidikan [2]. Pendidikan tinggi seperti STIKOM Dinamika Bangsa memiliki pangkalan data dari system informasi (sisfo) akademik. Saat ini pangkalan data telah menyimpan 7.174 record mahasiswa yang terdaftar semenjak tahun 2002. Pangkalan data sisfo saat ini telah menyimpan record mulai dari mahasiswa mengontrak mata kuliah tiap semester hingga wisuda.

Dengan tersedianya data sedemikian besar, biasanya terdapat informasi yang terpendam dan dapat digali, teknik yang dapat digunakan bisa menggunakan statistic deksriptif biasa hingga teknik Data Mining dan Exploratory Data Analysis (EDA) [3][4]. Informasi-informasi yang dapat kita temukan adalah pola-pola (pattern) serta korelasi antar variabel [5]. Jika ditemukan informasi-informasi tersebut tentunya akan berguna untuk dijadikan landasan untuk tahapan penelitian lebih lanjut seperti membangun model-model prediksi, pengelompokkan (clustering), dan rekomendasi.

Pada domain ini pernah dilakukan beberapa penelitian sejenis di antaranya Ferdiana, et al melakukan studi untuk mencari hubungan antara nilai Ujian Akhir Nasional(UAN) dengan Indeks Prestasi Kumulatif(IPK) dengan studi kasus di FMIPA Unsyiah [6], disimpulkan bahwa data yang didapat dari mahasiswa yang menempuh studi di FMIPA Unsyiah tersebut menunjukkan bahwa IPK berkorelasi positif dengan setiap nilai mata pelajaran Ujian Nasional, akan tetapi factor determinasinya akan cukup kecil (29,9% - 58,1%). Beta Norita dan Nurdi Bahtiar mencari hubungan antara tingkat kelulusan

mahasiswa dengan data induk mahasiswa. Data yang dipakai adalah data dari pangkalan data kampus UNDIP, data tersebut diambil dengan teknik data mining [7], Disimpulkan bahwa hubungan tingkat kelulusan dengan proses masuk mahasiswa jurusan matematika Universitas Diponegoro dengan threshold 0 untuk jalur Penjarangan Siswa-Siswa Berprestasi (PSSB) adalah dengan nilai support 0.8547, dan untuk siswa yang masuk melalui jalur Ujian Masuk Perguruan Tinggi (UMPTN) ternyata juga memiliki nilai support 0.8547. Dengan kata lain tingkat kelulusan dari dua jalur tersebut adalah sama.

Tujuan dari penelitian ini adalah mengidentifikasi variabel-variabel yang terdapat pada pangkalan data Sistem Informasi akademik STIKOM Dinamika Bangsa, serta mengidentifikasi pola-pola (pattern) dan relasi antar variabel yang telah diidentifikasi. Dengan tercapainya tujuan tersebut maka akan memudahkan penelitian berikutnya dalam mengukur seberapa kuat keterhubungan antar variabel tersebut.

## 2. METODE PENELITIAN

Dalam melakukan penelitian ini penulis melalui beberapa tahapan antara lain:

- a) Identifikasi Masalah  
Pada tahapan ini penulis melakukan identifikasi masalah yang berhubungan dengan sisfo akademik di STIKOM Dinamika Bangsa
- b) Pengumpulan dan Pengolahan Data  
Pada tahapan ini penulis mengumpulkan dan mengolah data yang didapat langsung dari pangkalan data sisfo akademik STIKOM Dinamika Bangsa
- c) Analisa Data  
Pada tahapan ini penulis menganalisa data yang sebelumnya melalui serangkaian proses terlebih dahulu, seperti pemilihan data, pembersihan data, hingga transformasi data.
- d) Penulisan Laporan  
Pada tahapan ini penulis membuat laporan akhir penelitian yang dibutuhkan.

## 3. PEMBAHASAN

Dengan melakukan visualisasi sebaran data, kita dengan cepat bisa mendapatkan gambaran normal tidaknya sebaran data tersebut. Di bagian ini kita melakukan visualisasi sebaran nilai ipk terhadap beberapa variabel dari data siswa.

### 3.1 Pengumpulan Data

Data berupa hasil dumping atau export seluruh data dari database Sistem Informasi Akademik milik STIKOM Dinamika Bangsa menjadi 2 file dengan tipe data comma-separated value (.csv). File pertama adalah gabungan dari beberapa table termasuk : table data diri siswa, table dosen, table matakuliah, table edom, dan table kontrak matakuliah. Data dari tabel-tabel tersebut kemudian dipisahkan ke file tersendiri.

Tabel 1.  
Kolom-kolom pada Tabel Siswa

No	Nama Kolom	No	Nama Kolom	No	Nama Kolom	No	Nama Kolom
1	nim	19	kd_dosen	37	diterima	55	telp_ot
2	nirl	20	pasfoto	38	prop_pt	56	pdk_ayah
3	nama	21	tmp_lahir	39	ijazah_pt	57	pdk_ibu
4	kls	22	tgl_lahir	40	jur_pt	58	krj_ayah
5	fak	23	kelamin	41	nim_asal	59	krj_ibu
6	jur	24	gol_darah	42	jml_sks	60	sts_ayah
7	angkatan	25	sts_sipil	43	nm_skul	61	sts_ibu
8	shift	26	no_ktp	44	sts_skul	62	anak_ke
9	gelombang	27	warga	45	jur_skul	63	jml_sdr
10	kurikulum	28	agama	46	thn_ijazah_skul	64	nm_pers
11	jng_studi	29	alamat	47	no_ijazah_skul	65	jabatan
12	sts_kul	30	kota	48	kabupaten	66	sts_login
13	tgl_masuk	31	kd_pos	49	provinsi_sekolah	67	sts_bayar

14	tgl_lulus	32	telp	50	nm_ayah	68	tgl_login
15	nomor_ijazah	33	email	51	nm_ibu	69	user_id
16	no_urut	34	sts_krj	52	alamat_ot	70	kelas
17	s_masuk	35	sts_tgl	53	kota_ot	71	password
18	bts_studi	36	sumber_biaya	54	kd_pos_ot	72	jenis_pt
						73	nama_pt

Tabel 2.  
Kolom-kolom pada Table Data Dosen

No	Nama Kolom	No	Nama Kolom
1	kd_dosen	8	email
2	nik	9	no_urut
3	nama	10	pasfoto
4	gelar_d	11	jenis
5	gelar_b	12	user_id
6	alamat	13	password
7	telp	14	tgl_login

Tabel 3.  
Kolom-kolom pada Table Data Matakuliah

No	Nama Kolom	No	Nama Kolom
1	kd_mk	4	sks
2	mt_kul	5	prasyarat
3	mt_kul_e	6	tampil

Tabel 4.  
Kolom-kolom pada table kontrak matakuliah

No	Nama Kolom	No	Nama Kolom
1	id	6	kontrak
2	nim	7	semester
3	kelas	8	periode
4	kd_mk	9	tgl_kontrak
5	sks		

Tabel 5.  
Kolom-kolom pada Table Hasil Edom

No	Nama Kolom
1	id
2	id_edom
3	kode_aspek
4	kode_kompetensi
5	score

Tabel 6.  
Kolom-kolom pada Tabel Edom

No	Nama Kolom
1	id_kompetensi
2	kode_kompetensi
3	jenis_kompetensi

### 3.2 Pemilihan Data

Dari beberapa table data yang didapat sebelumnya, tidak semuanya akan digunakan dalam penelitian ini. Dalam hal ini penulis memilih table data yang berhubungan dengan data siswa, sehingga pemillihan data jatuh kepada table siswa dan table indeks prestasi kumulatif (ipk).

Seperti yang dapat dilihat pada tabel di halaman sebelumnya, table data siswa banyak terdapat kolom-kolom yang tepatnya berjumlah 73 kolom. Kolom-kolom yang dipilih umumnya adalah kolom yang berjenis numeric, tanggal, dan kategori. Pertimbangan lain dalam memilih kolom yang dapat diolah adalah seberapa banyak nilai kosong yang terkandung di dalamnya. Adapun hasil pemilihan kolom yang penulis anggap relevan untuk penelitian ini berjumlah 18 kolom yang dapat dilihat pada tabel 3. Kolom Tabel Siswa Hasil Seleksi. Data berikutnya yang dipilih adalah data dari table indeks prestasi kumulatif (IPK). Tabel data ini hanya terdiri dari tiga (3) kolom yang dapat digunakan semuanya.

Tabel 2.  
Kolom-kolom pada Tabel IPK

No	Nama Kolom
1	NIM
2	Nama Mahasiswa
3	IPK

### 3.3 Pembersihan Data

Proses pembersihan data melibatkan penghapusan data ganda, pembenahan data, dan data scaling. Mengingat data yang didapat adalah berasal dari sebuah Database Management System yang telah terstruktur dengan cukup baik, penulis tidak menemukan data ganda pada data diri mahasiswa. Ditemukannya data mahasiswa yang sama pun dikarenakan mahasiswa tersebut pindah prodi, dan tetap tercatat sebagai mahasiswa pada prodi lama yang tidak menyelesaikan studi di prodi tersebut.

Tabel 2.  
Kolom Tabel Siswa Hasil Seleksi

No	Nama Kolom	No	Nama Kolom
1	nim	10	tmp_lahir
2	nama	11	tgl_lahir
3	jur	12	kelamin
4	angkatan	13	gol_darah
5	shift	14	agama
6	sts_kul	15	sts_krj
7	tgl_masuk	16	nm_skul
8	tgl_lulus	17	sts_skul
9	kd_dosen	18	jur_skul

Beberapa nama kolom yang didapat dari database dianggap kurang jelas dan kurang mudah dipahami atas apa yang akan menjadi isi dari kolom tersebut, oleh karena itu perubahan nama kolom penulis anggap perlu dilakukan. Nama kolom yang dirubah adalah:

- a) "jur" menjadi "jurusan",
- b) "sts\_kul" menjadi "status\_kuliah",

- c) "sts\_krj" menjadi "kerja",
- d) "nm\_skul" menjadi "sekolah",
- e) "sts\_skul" menjadi "status\_sekolah",
- f) "jur\_skul" menjadi "jurusan\_sekolah"

Dengan adanya perubahan nama kolom yang dianggap lebih bermakna tersebut, maka kolom-kolom untuk table data siswa yang telah diseleksi dan dinamai ulang, selain itu penulisan dengan huruf kecil semua juga dilakukan untuk tiap nama kolom, semua dapat dilihat pada table di bawah ini.

Tabel 3.  
Tabel Siswa Setelah Perubahan Nama Kolom

No	Nama Kolom	No	Nama Kolom
1	nim	10	tmp_lahir
2	nama	11	tgl_lahir
3	jurusan	12	kelamin
4	angkatan	13	gol_darah
5	shift	14	agama
6	status_kuliah	15	kerja
7	tgl_masuk	16	sekolah
8	tgl_lulus	17	status_sekolah
9	kd_dosen	18	jurusan_sekolah

Untuk data kosong yang umumnya bernilai NA umumnya akan diisi dengan nilai rata-rata atau nilai baru yang dianggap tidak akan mempengaruhi hasil pengolahan nanti. Kolom kd\_dosen berisi kode dosen yang merupakan dosen pembimbing akademik. Kolom ini terdapat data kosong yang kemudian diisi dengan nilai "xx". Hal tersebut dilakukan dengan tujuan data siswa tersebut akan tetap bisa digunakan jika nanti diperlukan perhitungan yang berhubungan dengan dosen. Untuk lebih lengkapnya pengisian data kosong dengan penggantinya adalah:

- a) kd\_dosen = "xx"
- b) tgl\_lulus = "tgl\_masuk"
- c) kerja = "belum bekerja"
- d) status\_sekolah = "negeri"
- e) jurusan\_sekolah = "ips"

Proses lain yang dilakukan pada tahapan ini adalah penggantian nilai pada beberapa kolom dari data siswa. Perubahan nilai pada kolom ini meliputi:

- a) Pada kolom jurusan berisi nilai 2, 3, atau 4. Nilai tersebut adalah kode untuk jurusan program studi TI, SI, dan SK.
- b) Pada kolom tgl\_masuk terdapat nilai 0000-00-00. Nilai tersebut dapat diisi dari nilai yang diambil dari kolom angkatan. Tanggal dan bulan diisi dengan tanggal 10 bulan 8. Nilai bulan cukup penting untuk diisi lebih akurat karena umumnya penerimaan siswa baru dimulai bulan 6 hingga 9.
- c) Kolom tgl\_keluar memiliki kasus serupa dengan kolom tgl\_masuk, penyebabnya adalah untuk mahasiswa yang tidak menyelesaikan kuliah baik drop out maupun tidak selesai. Untuk itu nilainya diisi sesuai dengan tanggal masuk sehingga lama kuliah nanti dapat dihitung.
- d) Kolom kerja terdapat tiga (3) nilai yaitu "Belum Bekerja", "Belum", dan "Sudah Bekerja". Nilai "Belum Bekerja" dan "Belum" adalah hal yang sama, oleh karena itu nilai "Belum" kita rubah menjadi "Belum Bekerja".

Data selanjutnya yang dipilih adalah dari table indeks prestasi kumulatif (IPK). Tidak ditemukan data kosong atau bernilai NA pada table data ini, sehingga perubahan yang diperlukan hanyalah pada penamaan kolom agar sesuai dengan nama kolom dari data mahasiswa. Ini diperlukan untuk memudahkan proses integrasi data nanti. Adapun perubahan nama kolom yang dilakukan adalah:

- a) "NIM" = "nim",
- b) "Nama.Mahasiswa" = "nama",
- c) "IPK" = "ipk"

### 3.4 Integrasi Data

Hasil dari pemilihan dan pembersihan data yang dilakukan pada proses sebelumnya kita gabungkan. Data mahasiswa dan ipk memiliki kolom bernilai unik yaitu kolom nim. Dari sini kita gabungkan data ipk dengan data mahasiswa berdasarkan nim. Sehingga data siswa akan memiliki kolom tambahan yaitu ipk. Dengan bertambahnya kolom ipk pada data siswa, kita dapat pemeriksaan silang antara status kuliah dengan nilai ipk. Ini dikarenakan mahasiswa yang memiliki nilai IPK adalah mahasiswa yang telah diwisuda. Sehingga siswa yang status kuliahnya tidak bernilai “Lulus” dapat kita koreksi menjadi “Lulus”.

### 3.5 Transformasi Data

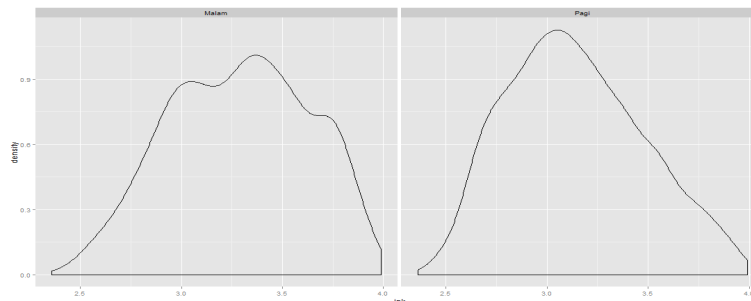
Dari data-data yang telah digabungkan menjadi satu pada proses integrasi data, tahapan selanjutnya adalah transformasi data. Ini diperlukan karena data-data yang dalam kondisi saat ini belum dapat menggambarkan kondisi yang ingin kita ketahui, terlebih untuk dicari hubungan linearnya, dalam hal ini adalah “persentase kelulusan”, dan “rata-rata ipk”. Hasil transformasi data didapatkan data baru seperti yang dapat dilihat pada table 3. Angkatan akan dibatasi hingga angkatan 2011 dengan alasan ketika data diambil, mahasiswa angkatan 2012 walaupun sudah memiliki nilai, persentasi kelulusan baru mencapai 37% dikarenakan ini adalah mahasiswa-mahasiswa yang lulus dengan durasi 3.5 tahun.

Tabel 3.  
Hasil Transformasi Data

angkatan	jumlah_lulusan	rata_rata	total_mhs	persentase_lulus
2002	189	3.088942	291	64.94845
2003	311	3.126206	514	60.50584
2004	349	3.105616	488	69.42675
2005	327	3.102752	471	69.42675
2006	332	3.140723	532	62.40602
2007	344	3.151017	494	69.63563
2008	383	3.147206	552	69.38406
2009	323	3.240588	478	67.57322
2010	308	3.235617	496	62.09677
2011	339	3.303245	568	59.6831
2012	208	3.459135	563	36.94494

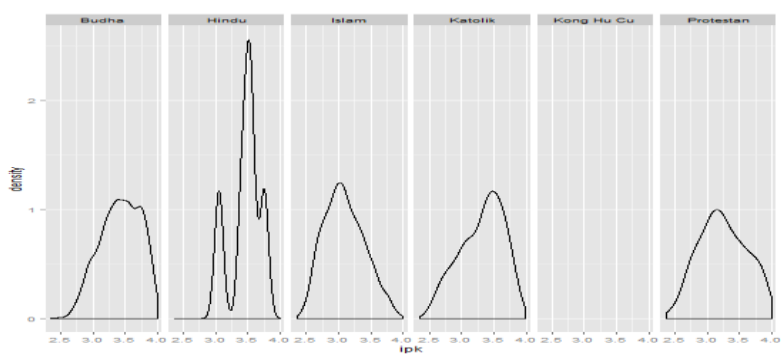
### 3.6 Visualisasi Sebaran Data

Dari data yang telah melalui proses sebelumnya, dapat kita diskusikan beberapa sebaran data ada.



Gambar 1. Kurva Densitas IPK terhadap shift kuliah malam atau pagi

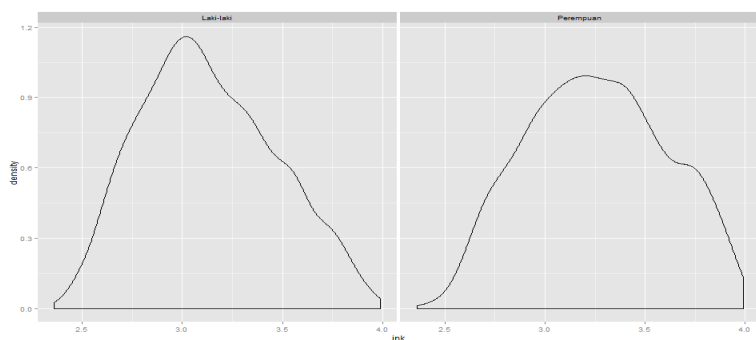
Terlihat bahwa bentuk kurva densitas sebaran nilai ipk pada mahasiswa yang dibedakan berdasarkan waktu kuliah malam masuk ke dalam kategori multimodal karena memiliki tiga (3) puncak. Ada kemungkinan bahwa mahasiswa/i di STIKOM Dinamika Bangsa yang mengambil jam kuliah malam terbagi menjadi tiga (3) kelompok. Mengingat ini adalah sebaran nilai ipk, bisa diprediksi tiga kelompok ini terbentuk berdasarkan pola belajar atau ada latar belakang lainnya. Untuk mahasiswa yang mengambil kelas pagi, seperti yang dapat kita lihat pada grafik, menunjukkan bentuk yang cenderung condong positif.



Gambar 2. Kurva densitas nilai IPK terhadap agama siswa

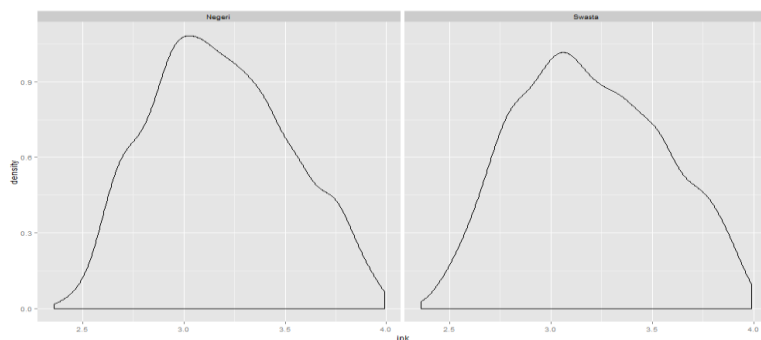
Seperti yang dapat dilihat pada Gambar 2, siswa dengan agama Hindu memiliki tiga (3) puncak dalam sebaran ipk yang lebih dikenal dengan istilah sebaran multimodal. Sebaran tidak normal ini dapat dikaji lebih lanjut mengenai penyebabnya.

Untuk agama lain seperti Budha, Islam, Katolik, memiliki sebaran yang cenderung miring. Siswa beragama Budha dan Katolik memiliki sebaran yang dengan kecondongan positif dimana puncak grafik berada di ipk bernilai di atas 3.0. Untuk mahasiswa beragama islam memiliki sebaran nilai ipk dengan kecondongan positif dengan titik grafik tertinggi berada di ipk di nilai 3.0. Mahasiswa dengan agama protestan memiliki sebaran yang cukup simetris dengan puncak grafik berada di atas ipk 3.0, sedangkan mahasiswa Bergama kong hu cu tidak dapat divisualkan sebaran nilai ipk-nya disebabkan frekuensi yang sedikit.



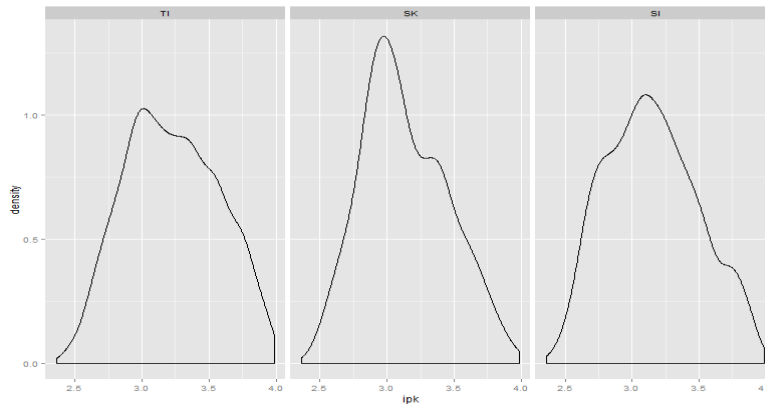
Gambar 3. Kurva densitas nilai ipk berdasarkan jenis kelamin

Puncak grafik nilai IPK pada mahasiswa dari sekolah swasta cenderung berada di ipk yang lebih tinggi dari anak dari sekolah negeri. Untuk sebaran nilai IPK untuk mahasiswa yang dibedakan berdasarkan jenis sekolah, yakni Negeri dan Swasta dapat dilihat pada gambar di bawah ini:



Gambar 4. Kurva densitas nilai IPK berdasarkan status sekolah

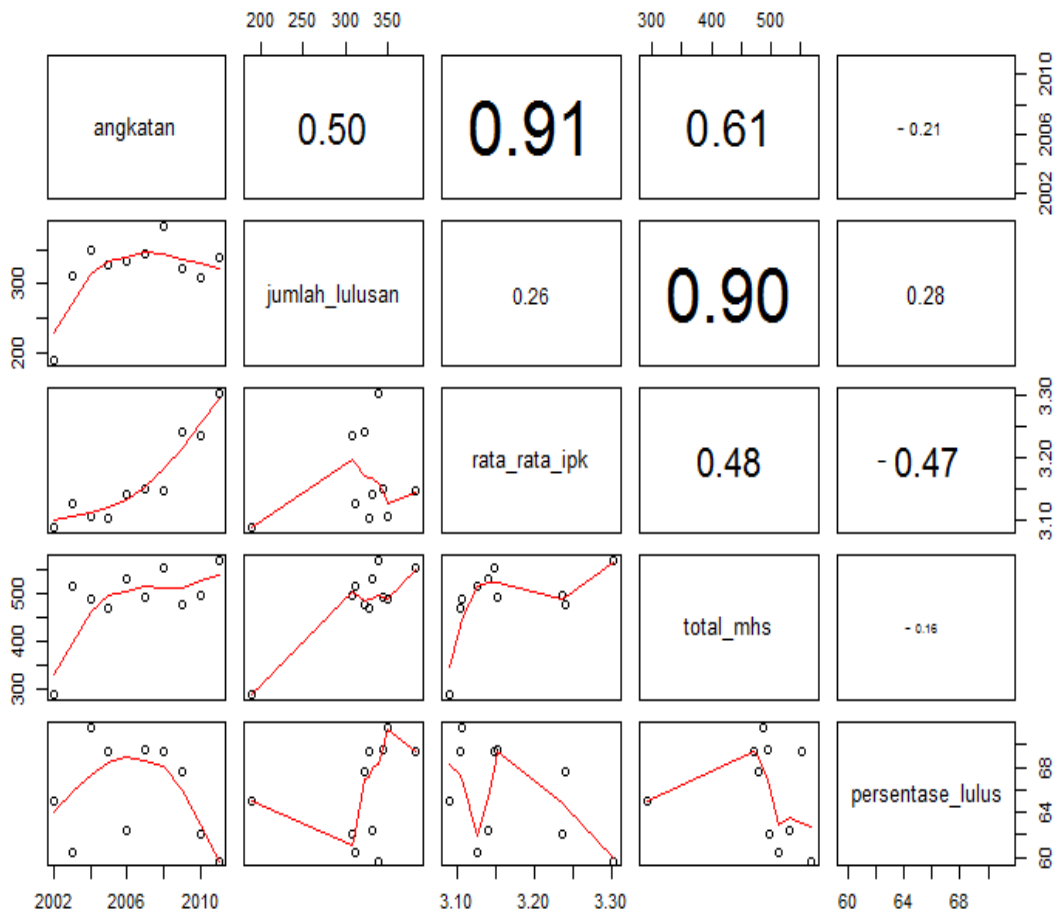
Ketika nilai ipk kita pisahkan berdasarkan jurusan, yakni Teknik Informatika (TI), Sistem Komputer (SK), dan Sistem Informasi (SI), terlihat kecenderungan terbentuknya kurva densitas multimodal (TI, SI) dan bimodal (SK).



Gambar 5. Kurva densitas nilai IPK berdasarkan jurusan

### 3.7 Analisa Korelasi Antar Variabel

Pada proses eksplorasi data yang mana kita gunakan metode visualisasi data, kita bisa dengan cepat melihat apakah beberapa variabel memiliki korelasi secara linear.



Gambar 2. Matrix Perbandingan Scatter Plot

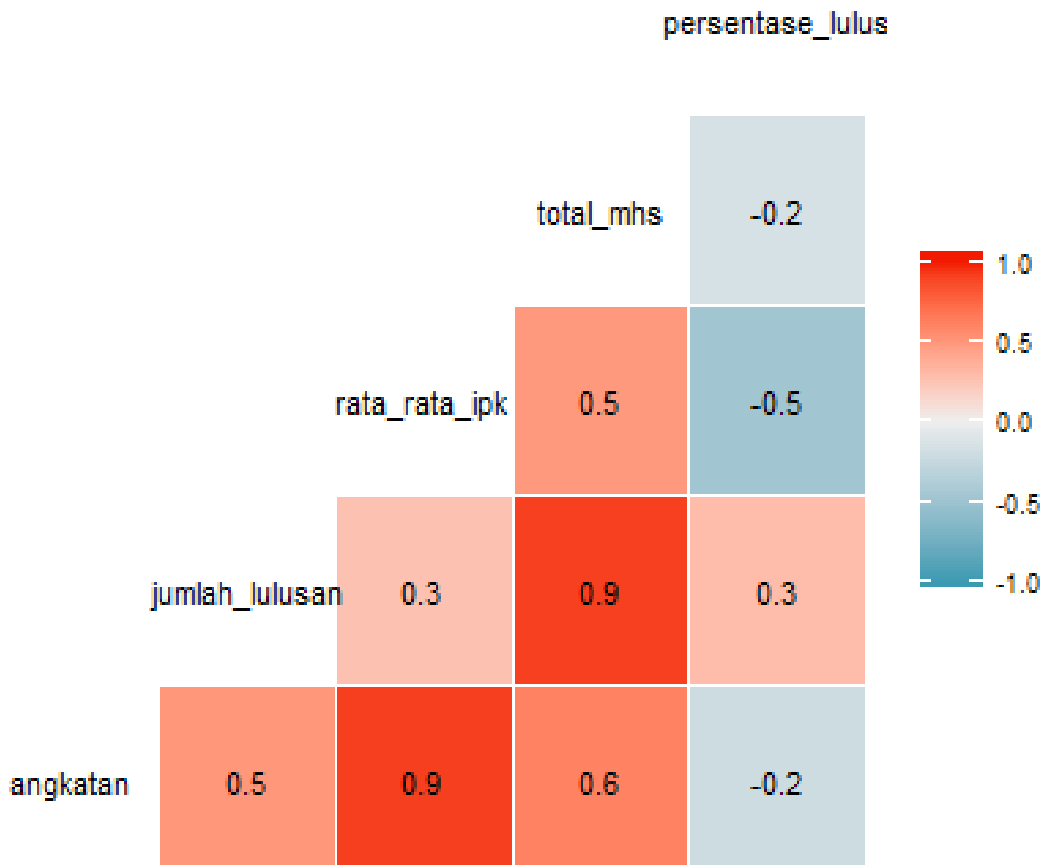


Variabel yang dipilih untuk dilihat korelasinya dalam hal ini adalah variabel yang telah didapat dari hasil transformasi data, yaitu:

- a) Angkatan
- b) Jumlah lulusan
- c) Rata-rata ipk
- d) Jumlah mahasiswa
- e) Persentase lulus

Terdapat tiga (3) hal yang bisa kita cermati:

- a) Variabel angkatan terhadap rata-rata ipk menunjukkan nilai korelasi Pearson Product Moment sebesar 0.91, dan pada Confidence Interval 95% didapatkan p-value sebesar 0.0003. Dalam grafik scatter plot terlihat garis linear positif, yang berarti tiap angkatan terjadi peningkatan nilai rata-rata ipk.
- b) Variabel rata-rata ipk terhadap persentase kelulusan menunjukkan nilai korelasi Pearson Product Moment sebesar -0.47, dan pada Confidence Interval 95% didapatkan p-value sebesar 0.1692. Dalam grafik menunjukkan adanya kecenderungan semakin banyak jumlah lulusan maka rata-rata ipk lulusan semakin kecil dengan semakin bertambahnya jumlah lulusan.
- c) Variabel total mahasiswa dengan jumlah lulusan menunjukkan nilai korelasi linear Pearson Product Moment sebesar 0.9, dan pada Confidence Interval 95% didapatkan p-value sebesar 0.0003. Hubungan ini dianggap terjadi cukup alami dikarenakan semakin banyak jumlah mahasiswa di tahun ajaran tersebut senantiasa semakin banyak juga jumlah lulusannya.



Gambar x. Matrix koefisien korelasi pearson product moment

#### 4. PENUTUP

Terdapat 18 variabel pada data diri mahasiswa yang teridentifikasi sebagai variabel yang mayoritas berisi nilai. Untuk melihat korelasi (linear) diperlukan proses transformasi data pada beberapa variabel yang memungkinkan terbentuknya variabel baru. Variabel baru ini antara lain rata-rata ipk, tingkat kelulusan, total mahasiswa. Variabel yang memiliki nilai koefisien korelasi Pearson Product Moment sebesar 0.9, pada Confidence Interval 95% dengan p-value 0.0003 adalah: angkatan terhadap ipk, total mahasiswa terhadap jumlah lulusan, variabel rata-rata ipk terhadap persentase kelulusan menunjukkan nilai koefisien korelasi Pearson Product Moment sebesar -0.47, pada Confidence Interval 95% dengan p-value 0.169.

Saran untuk penelitian berikutnya, dengan terbentuknya dataset siswa yang telah rapi (tidy data), memungkinkan untuk proses klusterisasi, prediksi, serta rekomendasi. Penelitian selanjutnya juga dapat mengarah ke model prediksi kapan mahasiswa akan mengambil cuti, non-aktif, maupun drop out.

#### DAFTAR PUSTAKA

- [1] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.
- [2] West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance Studies at Brookings*, 1-10.
- [3] O'Neil, Cathy and, Schutt, Rachel (2014) *Doing Data Science*, O'Reilly ISBN: 978-1-449-35865-5
- [4] Tukey, John W. The Future of Data Analysis. *Ann. Math. Statist.* 33 (1962), no. 1, 1--67. doi:10.1214/aoms/1177704711.
- [5] Heer, Jeffrey, Sean Kandel. 2012. "Interactive analysis of big data." *XRDS: Crossroads, The ACM Magazine for Students* 19.1 (): 50-54.
- [6] Ferdhiana Ridha, Ira Julita, Asep Rusyana, Nany Salwa, 2010 .“Hubungan Indeks Prestasi Kumulatif (IPK) dengan Nilai Ujian Akhir Nasional (UAN): Studi Kasus di FMIPA Unsyiah”, *Statistika* Vol. 15 N0.1, 17-23.
- [7] Beta Noranita, Nurdi Bahtiar. 2010. “Implementasi Data Mining Untuk Menemukan Pola Hubungan Tingkat Kelulusan Mahasiswa Dengan Data Induk Mahasiswa”, Seminar dan Call For Paper Munas Aptikom, Bandung.