



Integrasi Network Analysis dan Machine Learning Untuk Evaluasi Ekosistem Skripsi Mahasiswa Farmasi

Zakariyyah Hidayatulloh^{1*}, Esther Irawati Setiawan², Edwin Pramana³

^{1,2,3}Program Pascasarjana Magister Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya

E-Mail *: zakariyyah.h22@mhs.istts.ac.id

ABSTRACT

The thesis completion process is an absolute graduation requirement for undergraduate students, involving various stakeholders such as students, supervisors, and examiners. This study integrates Network Analysis and Machine Learning approaches to identify structural relationship patterns among lecturers, research topic trends, and predict student performance in the Pharmacy Undergraduate Study Program at Universitas Anwar Medika. The historical data used consists of 427 thesis records. Through the preprocessing stage, anomalous data on completion duration were eliminated, leaving valid data with an average completion time of 177.71 days and an average supervision frequency of 24.39 times. Network modeling produced a Degree Centrality matrix indicating the centralization of evaluation burdens on specific lecturers. The application of the Markov Clustering algorithm in Text Network Analysis of thesis titles revealed a polarization of research topics in the areas of pharmacognosy ("extract", "ethanol") and clinical pharmacy ("patient"). Furthermore, the K-Means Clustering algorithm successfully mapped three student academic performance profiles, and Linear Regression modeling proved capable of predicting final thesis grades with a very low Mean Absolute Error (MAE) of 1.60. The results of this study are expected to provide strategic insights for the university rectorate and study program administrators in distributing supervision workloads and directing the diversification of research topics in the future.

Keywords : Network Analysis, Centrality, Text Network Analysis, Skripsi

ABSTRAK

Proses penyelesaian skripsi merupakan syarat mutlak kelulusan bagi mahasiswa sarjana yang melibatkan berbagai pihak seperti mahasiswa, dosen pembimbing, dan penguji. Penelitian ini menerapkan pendekatan Network Analysis untuk mengidentifikasi pola hubungan struktural antar dosen serta tren topik penelitian mahasiswa di Program Studi S1 Farmasi Universitas Anwar Medika. Data yang digunakan berjumlah 427 rekaman data skripsi. Melalui tahap *preprocessing*, data anomali pada durasi penyelesaian dieliminasi, menyisakan data valid dengan rata-rata waktu penyelesaian 177,71 hari dan frekuensi bimbingan rata-rata 24,39 kali. Pemodelan jaringan menghasilkan matriks *Degree Centrality* yang menunjukkan dosen tertentu memiliki peran sentral yang dominan. Selain itu, *Text Network Analysis* pada judul skripsi mengungkapkan bahwa kata kunci "ekstrak", "pasien", dan "etanol" merupakan topik penelitian yang paling banyak memiliki interkoneksi. Hasil penelitian ini diharapkan dapat memberikan wawasan strategis bagi pihak rektorat dan program studi dalam mendistribusikan beban bimbingan serta mengarahkan variasi topik penelitian di masa mendatang.

Kata Kunci: Network Analysis, Centrality, Text Network Analysis, Skripsi



1. PENDAHULUAN

Skripsi merupakan salah satu syarat kelulusan bagi mahasiswa program studi S1 Farmasi Universitas Anwar Medika. Proses penyelesaian skripsi membutuhkan waktu dan tenaga yang cukup besar, melibatkan interaksi intensif antara mahasiswa sebagai peneliti, dosen pembimbing sebagai mentor, dan penguji sebagai penilai. Analisis hubungan interaksi antar civitas akademika sangat krusial untuk menemukan strategi guna meningkatkan efisiensi dan efektivitas proses penyelesaian skripsi.

Network analysis (analisis jaringan) merupakan salah satu metode analisis data yang mumpuni untuk menganalisis struktur dan hubungan antar entitas dalam suatu ekosistem yang kompleks. Dalam penerapannya di dunia pendidikan, metode analisis jaringan dapat memberikan informasi tentang kekuatan hubungan antara mahasiswa, dosen pembimbing, dan penguji. Beberapa studi terdahulu telah membuktikan efektivitas analisis jaringan dan *machine learning* dalam bidang akademik. Penelitian Tuhuteru & Iriani (2018) memanfaatkan *Social Network Analysis* (SNA) untuk menganalisis kolaborasi dosen, sementara Saqr & López-Pernas (2021) membuktikan pentingnya pemetaan dinamika temporal jaringan sosial mahasiswa. Pada evaluasi performa, Crossette dkk. (2023) menunjukkan korelasi antara konektivitas jaringan dengan kinerja akademik, sejalan dengan temuan Rastrollo-Guerrero dkk. (2020) dan Alyahyan & Düstegör (2020) bahwa *educational data mining* mampu mendeteksi faktor penentu kelulusan. Sementara itu, pada ranah pemrosesan teks, *Text Network Analysis* telah berhasil memetakan tren teknologi pendidikan (Chen dkk., 2020), isu strategis berbasis dokumen (Allipour Birgani dkk., 2022), hingga analisis tren media (Zhang dkk., 2023; Zhao & Li, 2023). Kebutuhan institusi untuk bertransformasi menuju metrik pengukuran performa yang presisi dan objektif juga telah dibuktikan urgensinya melalui penerapan sistem komputasi evaluasi kinerja berbasis data (Diana dkk., 2017).

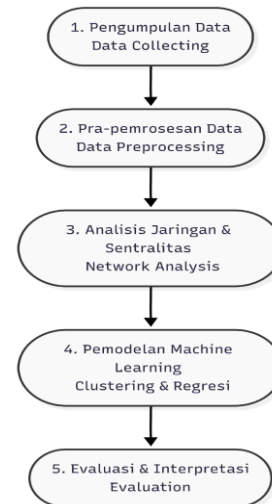
Meskipun berbagai studi telah dilakukan, terdapat kesenjangan penelitian (*research gap*) yang signifikan. Mayoritas penelitian sebelumnya cenderung mengisolasi metode; hanya berfokus pada analisis interaksi sosial (SNA) saja atau murni pada penambahan teks (*Text Mining*) saja, tanpa mengaitkannya dengan prediksi performa evaluasi akhir. Kebaruan (*novelty*) dari penelitian ini terletak pada pendekatan analitik hibrida yang secara simultan mengintegrasikan pemodelan struktural (*One-Mode Network*), pemetaan semantik (*Text Network Analysis*), dan algoritma *Machine Learning* (*K-Means* & Regresi Linear) dalam satu kerangka kerja utuh untuk membedah ekosistem penyelesaian tugas akhir.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk: (1) memetakan topologi jaringan interaksi struktural antar dosen pembimbing dan penguji, (2) mendeteksi klusterisasi tren topik riset mahasiswa secara *unsupervised*, dan (3) memvalidasi

serta memprediksi korelasi antara interaksi bimbingan, rekam jejak akademik, dan capaian nilai akhir skripsi mahasiswa menggunakan *Machine Learning*.

2. METODE

2.1 Kerangka Kerja Usulan (*Proposed Framework*)



Gambar 1. Frameworks Penelitian

Kerangka kerja penelitian yang diusulkan terdiri dari lima tahapan utama yang berjalan secara sekuensial:

1. Pengumpulan Data (*Data Collecting*): Ekstraksi data mentah skripsi S1 Farmasi dari pangkalan data Universitas Anwar Medika.
2. Pra-pemrosesan Data (*Data Preprocessing*): Pembersihan data dari anomali, penanganan data kosong (*missing values*), serta pemfilteran kata-kata umum (*stopwords*) untuk pengolahan teks.
3. Analisis Jaringan dan Sentralitas (*Network Analysis*): Visualisasi graf (*node* dan *edge*) dilanjutkan dengan kalkulasi metrik *Centrality* serta *Text Network Analysis* pada judul riset.
4. Pemodelan *Machine Learning*: Penerapan algoritma *K-Means Clustering* untuk pemetaan profil akademik dan Regresi Linear untuk prediksi probabilitas nilai akhir.
5. Evaluasi dan Interpretasi (*Evaluation*): Pengukuran validitas model (*Silhouette Score*, MAE, MSE) dan penarikan kesimpulan strategis.

Alur penelitian ini disusun secara sistematis agar identifikasi pola jaringan dapat dilakukan dengan terstruktur.

1. Identifikasi Masalah: Memahami kendala dalam penyelesaian skripsi di lingkungan program studi.
2. Studi Literatur: Mempelajari analisis data yang akan digunakan untuk penelitian ini.
3. Pengumpulan dan Pengolahan Data: Data yang akan digunakan dalam penelitian ini mencakup informasi tentang skripsi yang telah selesai pada

- program studi S1 Farmasi. Pembersihan data dilakukan dari *noise*, duplikasi, data kosong, atau kesalahan. Transformasi dilakukan ke format yang sesuai untuk ekstraksi pola komputasional.
4. Penyusunan Algoritma: Algoritma berbasis bahasa *Python* digunakan untuk memetakan *Network Analysis* dan menganalisa data dari *database*.
 5. Simpulan Hasil: Hasil eksperimen akan diinterpretasikan untuk memberikan makna terhadap data dan menjawab pertanyaan penelitian. Visualisasi data digunakan untuk menyajikan hasil analisis data

2.2 Dataset

Dataset yang digunakan merupakan data historis sekunder yang diekstraksi dari pangkalan data akademik. Karakteristik variabel penelitian diuraikan pada Tabel 1.

Tabel 1. Tabel

Atribut	Tipe Data	Deskripsi	Pemodelan
Dosen Pembimbing	Kategori	Identitas anonim dosen pembimbing utama	Node (<i>One-Mode Network</i>)
Dosen Penguji	Kategori	Identitas anonim dosen penguji utama	Node (<i>One-Mode Network</i>)
Judul Skripsi	Teks	String teks judul penelitian mahasiswa	Ekstraksi Node (<i>Text Network</i>)
Waktu Penyelesaian	Numerik	Durasi pengerjaan skripsi (dalam hari)	Fitur Prediksi (<i>Machine Learning</i>)
Jumlah Bimbingan	Numerik	Frekuensi tatap muka bimbingan	Fitur Clustering (<i>Machine Learning</i>)
IPK	Numerik	Indeks Prestasi Kumulatif mahasiswa	Variabel Independen (Regresi)
Nilai Skripsi	Numerik	Nilai akhir kelulusan sidang	Variabel Dependen / Target

2.3 Knowledge Discovery in Databases (KDD) dan Data Mining

Knowledge Discovery in Databases merupakan proses menyeluruh dalam menggali serta menganalisis data untuk menemukan informasi yang bernilai. Tahapan ini mencakup seleksi data, pembersihan (*preprocessing*), transformasi data, *data mining*, dan interpretasi atau evaluasi. *Data mining* sendiri adalah suatu proses untuk menemukan sebuah pola, pengetahuan, atau informasi yang berguna sebagai pendukung keputusan yang lebih baik dari sejumlah besar data (Han dkk., 2012).

2.4 Network Analysis

Analisis jaringan (*network analysis*) adalah suatu pendekatan ilmiah dan metodologis yang digunakan untuk memodelkan, memvisualisasikan, dan memahami struktur, pola, serta interaksi dalam suatu sistem yang kompleks. Ekosistem jaringan direpresentasikan melalui sebuah graf yang secara mendasar terdiri dari dua komponen utama, yaitu simpul (*node* atau *vertex*) dan garis penghubung (*edge* atau *link*) (Pavlopoulos dkk., 2017; Setiawan dkk., 2016).

- *Node* (Simpul): Merupakan entitas diskrit di dalam jaringan. Dalam konteks akademik, *node* dapat merepresentasikan individu seperti mahasiswa dan dosen, atau dalam konteks tekstual, *node* dapat merepresentasikan sebuah kata atau konsep.
- *Edge* (Garis Penghubung): Merupakan representasi dari relasi atau interaksi yang menghubungkan antar *node*. *Edge* dapat bersifat *directed* (berarah, misalnya hierarki bimbingan) maupun *undirected* (tidak berarah, misalnya relasi kolaborasi pengujian dalam satu sidang).

Secara arsitektur, pemodelan jaringan dalam penelitian ini memanfaatkan dua pendekatan spesifik:

- *One-Mode Network* (Jaringan Satu Arah): Merupakan jenis pemodelan jaringan di mana seluruh *node* yang terlibat berasal dari himpunan entitas yang seragam (satu tipe). Dalam penelitian ini, *one-mode network* digunakan untuk memetakan hubungan sosial antar sesama dosen (Pembimbing dan Penguji) yang dihubungkan oleh aktivitas akademik yang sama, sehingga memungkinkan identifikasi komunitas (*clustering*) atau kelompok keilmuan secara organik.
- *Text Network Analysis* (TNA): Merupakan perpaduan antara linguistik komputasional dan analisis jaringan. TNA tidak menganalisis entitas manusia, melainkan memetakan kata kunci dari dokumen (seperti judul skripsi) sebagai *node*. *Edge* dalam TNA terbentuk berdasarkan *co-occurrence* (kemunculan bersama) dari dua kata dalam rentang konteks yang sama. Metode ini memungkinkan ekstraksi arsitektur pengetahuan dan deteksi tren topik yang tidak bisa ditangkap oleh analisis frekuensi

kata konvensional.

2.5 Sentralitas (Centrality)

Centrality adalah metrik algoritma yang krusial dalam *network analysis* untuk mengidentifikasi dan mengukur tingkat kepentingan, pengaruh, atau dominasi suatu node di dalam keseluruhan arsitektur jaringan (Javed dkk., 2021). Terdapat tiga parameter pengukuran sentralitas yang digunakan secara komprehensif dalam penelitian ini, yang sejalan dengan kerangka evaluasi struktural kontemporer (Bail, 2021):

1. *Degree Centrality*: Metrik dasar yang mengukur jumlah koneksi langsung (insiden) yang dimiliki oleh sebuah *node*. Semakin tinggi nilai *degree*, semakin aktif *node* tersebut berinteraksi dalam jaringan. Perhitungannya dirumuskan dengan:

$$C_D(v_i) = d(v_i) \quad (1)$$

di mana $d(v_i)$ melambangkan derajat atau jumlah garis yang terhubung pada *node* v_i .

2. *Closeness Centrality*: Metrik ini mengukur tingkat kedekatan suatu *node* terhadap seluruh *node* lain dalam jaringan berdasarkan kalkulasi jarak jalur terpendek (*shortest path*). *Node* dengan *closeness* tinggi memiliki keunggulan aksesibilitas, menjadikannya penyebar informasi yang paling efisien. Rumus yang digunakan adalah:

$$C_c(v_i) = \frac{N-1}{\sum_j d(v_i, v_j)} \quad (2)$$

di mana N adalah jumlah total *node* dalam jaringan, dan $d(v_i, v_j)$ adalah jarak geodesik (*geodesic distance*) dari *node* i ke *node* j .

3. *Betweenness Centrality*: Mengukur seberapa sering sebuah *node* bertindak sebagai "jembatan" (*broker/bridge*) yang menghubungkan *node-node* lain yang tidak saling terhubung secara langsung. Entitas dengan *betweenness* tinggi memegang kontrol kritis terhadap aliran informasi lintas komunitas. Formulasi matematisnya adalah:

$$C_B(v_i) = \sum_{j \neq k \neq i} \frac{g_{jk}(v_i)}{g_{jk}} \quad (3)$$

di mana g_{jk} merepresentasikan total jumlah jalur terpendek dari *node* j ke k , dan $g_{jk}(v_i)$ merupakan fraksi dari jalur terpendek tersebut yang harus melewati *node* i .

2.6 Komposisi Node dan Formalisasi Pembentukan Edge

Untuk memberikan kejelasan metodologis, representasi objek graf di dalam penelitian ini secara formal dibagi menjadi dua arsitektur jaringan terpisah

dengan definisi pembentukan garis penghubung (*edge*) sebagai berikut:

1. Jaringan Interaksi Dosen (*One-Mode Network*)
 - *Node (V)*: Merepresentasikan entitas dosen di lingkungan Program Studi S1 Farmasi Universitas Anwar Medika, tanpa membedakan status fungsional secara permanen.
 - *Edge (E)*: Hubungan (*link*) dibentuk secara tidak berarah (*undirected*) berlandaskan relasi Pembimbing Utama – Penguji Utama. Sebuah *edge* akan tercipta di antara dua dosen jika dan hanya jika Dosen A (sebagai Pembimbing Utama) dan Dosen B (sebagai Penguji Utama) ditugaskan bersama dalam satu majelis sidang skripsi mahasiswa yang sama. Akumulasi dari kedekatan tugas yang berulang pada mahasiswa yang berbeda akan dikalkulasi sebagai bobot (*weight*) dari *edge* tersebut. Jaringan ini secara sengaja mengeliminasi relasi Pembimbing–Mahasiswa maupun Penguji–Mahasiswa untuk berfokus murni pada pemetaan struktur kolaborasi, kemitraan evaluasi, dan distribusi beban kerja di tingkat pendidik.
2. Jaringan Konten Riset (*Text Network Analysis*)
 - *Node (V)*: Merepresentasikan kata kunci (*keywords*) tunggal yang berhasil diekstrak dari dokumen judul skripsi mahasiswa.
 - *Edge (E)*: Hubungan dibentuk berdasarkan metrik *co-occurrence* (kemunculan bersama). Sebuah *edge* tidak berarah (*undirected*) akan ditarik menghubungkan Kata X dan Kata Y jika kedua kata tersebut muncul secara bersamaan di dalam satu judul skripsi yang sama.

2.7 Pra-pemrosesan Data (Preprocessing)

Proses *preprocessing* data krusial untuk memastikan validitas hasil komputasi algoritma. Dalam penelitian ini, analisis pendahuluan menemukan data anomali ekstrem (waktu penyelesaian bernilai negatif atau melebihi 1000 hari). *Noise* ini direduksi melalui *data cleaning*. Setelah data disaring (waktu penyelesaian antara 30 hingga 1000 hari), diperoleh data yang merepresentasikan tren normal penyelesaian skripsi. Pada data tekstual (Judul Skripsi), dilakukan penghapusan tanda baca, konversi teks menjadi huruf kecil, serta eliminasi *stopwords* bahasa Indonesia untuk mendapatkan ekstraksi kata kunci murni.

2.8 Perangkat Lunak dan Lingkungan Komputasi

Seluruh tahapan manipulasi data, pemodelan struktur topologi graf, kalkulasi matematis metrik sentralitas, hingga eksekusi algoritma deteksi komunitas dilakukan secara terprogram

memanfaatkan ekosistem bahasa pemrograman Python (Versi 3). Beberapa pustaka (*open-source libraries*) spesifik yang diintegrasikan sebagai instrumen komputasi dalam penelitian ini meliputi:

1. Pandas: Berperan dalam proses *data ingestion* dari pangkalan data administrasi skripsi, penanganan data kosong (*missing values*), serta penyaringan data anomali durasi penyelesaian secara tabular.
2. NetworkX: Merupakan *software engine* utama yang digunakan untuk mengonversi data relasional menjadi struktur graf, menghitung matriks sentralitas (*Degree, Closeness, Betweenness Centrality*), serta mengeksekusi algoritma *modularity-based community detection* untuk memetakan kluster dosen dan topik secara otomatis.
3. Matplotlib: Digunakan sebagai pustaka visualisasi dasar untuk merender tata letak (*layout*) jaringan, visualisasi sebaran simpul, dan pemetaan garis interaksi secara langsung dari objek graf yang dibangun oleh NetworkX.

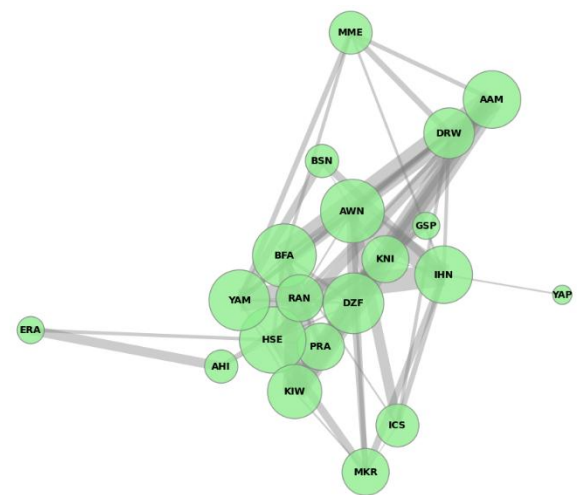
3 PEMBAHASAN

3.1 Analisis Sentralitas Jaringan Dosen (*One-Mode Network*)

Jaringan satu arah (*one-mode network*) dibangun berdasarkan relasi antara Pembimbing Utama dan Penguji dalam satu sidang. Tingginya sentralitas pada dosen-dosen tertentu, seperti HSE dan AWN tidak hanya sekadar mendeskripsikan beban bimbingan atau pengujian yang tinggi, tetapi mengindikasikan adanya pemusatan kepakaran spesifik. Ketimpangan beban struktural ini secara analitik kemungkinan besar didorong oleh dua faktor utama: pertama, rasio ketersediaan dosen ahli di bidang peminatan spesifik (seperti Farmasi Klinis) belum sebanding dengan tingginya animo mahasiswa di bidang tersebut. Kedua, adanya integrasi peran struktural seperti Kepala Laboratorium yang menuntut dosen bersangkutan untuk selalu dilibatkan sebagai verifikator kelayakan teknis riset, sehingga otomatis menjadikan mereka *hub* utama dalam jaringan evaluasi.

Melalui pemeringkatan metrik *Degree Centrality*, diidentifikasi tiga nama *node* (dosen) yang memiliki nilai pengaruh koneksi tertinggi:

1. HSE
2. AWN
3. BFA



Gambar 2. Degree Centrality Dosen

Hasil ini mengindikasikan bahwa ketiga dosen tersebut sering kali berurusan tugas sebagai penguji maupun pembimbing satu sama lain, sehingga memegang peranan krusial sebagai jembatan informasi akademik di prodi Farmasi.

Selain pengukuran metrik *Degree Centrality*, penelitian ini juga mengevaluasi *Closeness* dan *Betweenness Centrality* untuk memahami secara mendalam peran struktural dari masing-masing dosen di dalam jaringan akademik. Nilai dari ketiga parameter sentralitas ini dapat dilihat pada Tabel 2 di bawah ini.

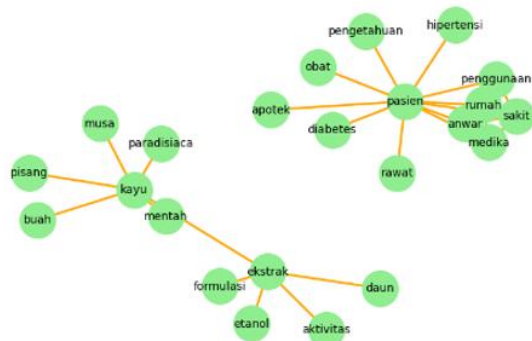
Tabel 2. Tabel Centrality Dosen

Dosen	Degree	Closeness	Betweenness
HSE	0.6316	0.7037	0.2449
AWN	0.5789	0.7037	0.0950
BFA	0.5789	0.6786	0.0719
YA	0.5263	0.6786	0.0957
DZF	0.5263	0.6552	0.0420
IFN	0.4737	0.6129	0.1572
AAM	0.4737	0.6333	0.0399
KIW	0.4211	0.6129	0.0161
DRA	0.3684	0.5588	0.0147

Berdasarkan Tabel 2, dapat diamati bahwa dosen dengan nilai *Degree Centrality* tertinggi cenderung juga memiliki nilai *Closeness Centrality* yang signifikan. Nilai *Closeness* yang tinggi menandakan bahwa dosen tersebut memiliki jarak interaksi yang sangat dekat dengan mayoritas pendidik lainnya dalam jaringan. Hal ini menjadikan mereka figur yang sangat efisien dalam sirkulasi informasi akademik antar dosen.

Di sisi lain, metrik *Betweenness Centrality* menyoroti dosen-dosen yang bertindak sebagai "jembatan" atau *broker* di antara kelompok-kelompok dosen yang mungkin jarang berinteraksi satu sama lain. Dosen dengan skor *Betweenness*

tertinggi seperti HSE (0.2449) dan IHN (0.1572) memegang kontrol kritis terhadap aliran informasi, standar evaluasi, serta distribusi kebijakan skripsi antar berbagai klaster peminatan di Program Studi S1 Farmasi Universitas Anwar Medika.



Gambar 3. Visualisasi *Text Network Analysis*

3.2 Text Network Analysis

Pada Judul Skripsi Pendekatan *Text Network Analysis* memetakan relasi antar-kata kunci (*co-occurrence*) pada seluruh judul skripsi mahasiswa. Berdasarkan perhitungan *Centrality*, didapatkan 5 kata kunci paling sentral:

1. ekstrak (Skor Centrality: 0.3523)
2. pasien (Skor Centrality: 0.3008)
3. etanol (Skor Centrality: 0.2455)
4. aktivitas (Skor Centrality: 0.2295)
5. metode (Skor Centrality: 0.2205)

Hal ini menunjukkan bahwa riset mahasiswa sarjana Farmasi Universitas Anwar Medika memiliki konsentrasi masif pada studi mengenai metode ekstraksi senyawa berpelarut etanol untuk menguji aktivitas biologis, serta penelitian berbasis kepuasan dan pelayanan klinis terhadap pasien.

Dominasi kata kunci 'ekstrak', 'etanol', dan 'pasien' sebagai *node* paling sentral mengungkapkan bahwa riset mahasiswa S1 secara konsisten mengumpul pada dua pilar: Farmakognosi (bahan alam) dan Farmasi Klinis. Pemusatan tren pada ekstraksi berpelarut etanol ini tidak terjadi secara kebetulan, melainkan dipengaruhi kuat oleh ketersediaan instrumen laboratorium ekstraksi di Universitas Anwar Medika yang memadai, serta biaya operasional riset bahan alam yang relatif lebih terukur dan mudah dijangkau oleh mahasiswa sarjana dibandingkan riset sintesis obat. Sementara itu, tingginya interkoneksi kata 'pasien' mencerminkan arah kebijakan kurikulum program studi yang saat ini gencar mendorong penelitian terapan. Riset berorientasi pelayanan klinis rumah sakit lebih banyak diminati karena kemudahan akses data rekam medis melalui kolaborasi institusi dan selaras dengan kompetensi profil lulusan yang dicanangkan.

3.3 Korelasi Interaksi Bimbingan dan Performa Akademik

Terkait performa akademik, hasil uji korelasi Pearson menunjukkan terdapat hubungan positif linear yang signifikan secara statistik antara Indeks Prestasi Kumulatif (IPK) dan Nilai Skripsi ($r = 0,445$; $p\text{-value} < 0,001$). Uji signifikansi ini didukung oleh rentang *Confidence Interval* (CI) 95% yang berada pada $[0,366; 0,519]$. Mengingat $p\text{-value}$ yang jauh di bawah ambang batas signifikansi 0,05 dan rentang CI yang tidak melewati angka nol, dapat divalidasi bahwa hubungan antara rekam jejak akademik (IPK) dengan hasil evaluasi akhir (Nilai Skripsi) benar-benar nyata, valid, dan tidak terjadi secara kebetulan. Hal ini sejalan dengan penelitian (Crossette dkk., 2023), yang menyimpulkan bahwa konektivitas akademik mahasiswa beririsan langsung dengan hasil performa akhir.

3.4 Pemetaan Klaster Topik Riset Skripsi

Lebih lanjut, metode algoritma deteksi komunitas juga diterapkan pada jaringan *co-occurrence* (kemunculan bersama) kata kunci judul skripsi mahasiswa. Dari hasil pemetaan *Text Network Analysis*, ditemukan dua klaster komunitas kata kunci yang paling mendominasi dan memiliki bobot koneksi tertinggi:

1. Komunitas Riset Eksperimental Laboratorium: Kata kunci yang saling berkumpul membentuk klaster ini meliputi: "ekstrak", "aktivitas", "etanol", "metode", "formulasi", "kayu", "daun", dan "buah". Komunitas ini mewakili tren dominan mahasiswa yang mengambil skripsi berbasis pengujian *in vitro*, optimasi formulasi sediaan farmasi, serta ekstraksi metabolit sekunder dari bahan alam.
2. Komunitas Riset Farmasi Rumah Sakit/Klinis: Terbentuk dari himpunan kata kunci: "pasien", "penggunaan", "rumah", "obat", "sakit", "diabetes", dan "medika". Komunitas ini menegaskan adanya ekosistem penelitian observasional yang sangat kuat yang berfokus pada evaluasi penggunaan obat, tingkat kepatuhan, serta kepuasan pelayanan pasien dengan mengambil studi kasus langsung di Rumah Sakit Anwar Medika.

Penemuan klaster ini memberikan *novelty* temuan bahwa tren riset mahasiswa farmasi telah terpolarisasi secara spesifik ke dalam dua kutub utama. Wawasan ini dapat menjadi pijakan bagi manajemen program studi untuk menyiapkan alokasi dana laboratorium dan perizinan etika klinis rumah sakit yang lebih proporsional di masa mendatang.

3.5 Clustering Profil Akademik Mahasiswa

Selain pendekatan *Network Analysis*, penelitian ini juga menerapkan algoritma *Machine Learning unsupervised*, yaitu *K-Means Clustering*, yang

terbukti tangguh dalam mempartisi data multidimensi secara otomatis untuk memetakan profil performa akademik mahasiswa, sebagaimana diadaptasi dalam kerangka presisi pendidikan mutakhir (Sarker, 2021; Luan & Tsai, 2021; Ouyang & Jiao, 2021). Parameter yang digunakan untuk pengelompokan meliputi Indeks Prestasi Kumulatif (IPK), Nilai Skripsi, dan Waktu Penyelesaian. Berdasarkan hasil komputasi, terbentuk 3 (tiga) klaster profil mahasiswa:

1. Klaster 1 (Performa Sedang): Kelompok mayoritas dengan rata-rata IPK 3,16, Nilai Skripsi 81,95, dan waktu penyelesaian 179 hari.
2. Klaster 2 (Performa Tinggi): Kelompok unggulan dengan rata-rata IPK tertinggi (3,45), Nilai Skripsi sangat baik (83,11), dan waktu penyelesaian paling cepat (168 hari).
3. Klaster 3 (Performa Bawah): Kelompok dengan rata-rata IPK terendah (2,97), Nilai Skripsi standar (77,42), dan waktu penyelesaian terlama (188 hari).

Untuk mengevaluasi kualitas dan tingkat keakuratan dari *clustering* yang dihasilkan, dilakukan perhitungan nilai error. Pemodelan ini menghasilkan nilai *Sum of Squared Error* (SSE/Inertia) sebesar 237,67. Jika dikonversi menjadi rata-rata kesalahan per titik data terhadap pusat klasternya (*Mean Squared Error* / MSE untuk *clustering*), model ini memiliki skor MSE sebesar 1,60 dan nilai *Silhouette Score* sebesar 0,30. Angka *error* yang rendah ini mengindikasikan bahwa pembagian profil kelulusan mahasiswa ke dalam tiga klaster tersebut cukup padat dan representatif.

3.6 Pemodelan Prediksi Nilai Skripsi Mahasiswa

Berdasarkan korelasi positif yang ditemukan antara rekam jejak akademik dan evaluasi akhir, penelitian ini lebih lanjut membangun model *supervised learning* (Regresi Linear) untuk memprediksi probabilitas Nilai Skripsi berdasarkan data IPK dan Durasi Penyelesaian. Hal ini mengadaptasi kerangka prediksi performa mahasiswa berbasis *machine learning* yang terbukti efektif menekan tingkat galat evaluasi (Hasan dkk., 2021; Namoun & Alshanqiti, 2021; Yağcı, 2022; Bogarín dkk., 2018).

Kinerja algoritma prediksi ini dievaluasi secara matematis menggunakan dua metrik tingkat kesalahan baku, yaitu *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Hasil pengujian model pada dataset menghasilkan skor MSE sebesar 4,77 dan skor MAE sebesar 1,60. Nilai MAE yang hanya berada di angka 1,60 poin membuktikan bahwa penyimpangan (selisih) rata-rata antara nilai skripsi hasil prediksi model dengan nilai asli mahasiswa sangatlah kecil. Hal ini menegaskan bahwa IPK dan manajemen waktu penyelesaian dapat dijadikan prediktor yang cukup baik terhadap hasil akhir nilai sidang skripsi di Program Studi S1 Farmasi.

4 KESIMPULAN

Penelitian ini telah berhasil mengintegrasikan pendekatan Network Analysis dan Machine Learning untuk menggali wawasan komprehensif dari data administrasi skripsi Program Studi S1 Farmasi Universitas Anwar Medika. Berdasarkan analisis sentralitas dan deteksi komunitas (Community Detection), terungkap bahwa struktur interaksi dosen secara organik membentuk tiga klaster keilmuan dengan pemusatan peran evaluasi pada figur-figur sentral tertentu. Selain itu, pemetaan Text Network Analysis menegaskan adanya polarisasi tren riset mahasiswa ke dalam dua kutub utama, yaitu riset eksperimental laboratorium (farmakognosi) dan riset farmasi klinis.

Pada aspek performa akademik, hasil uji statistik mengonfirmasi korelasi positif yang signifikan antara rekam jejak akademik (IPK) dan nilai evaluasi akhir skripsi. Pemodelan unsupervised learning melalui K-Means Clustering secara akurat berhasil memetakan mahasiswa ke dalam tiga profil kelulusan. Lebih lanjut, pemanfaatan algoritma Regresi Linear terbukti mampu memprediksi probabilitas nilai akhir skripsi mahasiswa berdasarkan parameter IPK dan waktu penyelesaian dengan tingkat penyimpangan yang sangat rendah, yang dibuktikan dengan skor Mean Absolute Error (MAE) sebesar 1,60 dan Mean Squared Error (MSE) sebesar 4,77.

Integrasi hasil penelitian ini diharapkan dapat menjadi landasan strategis bagi pihak rektorat dan program studi. Wawasan ini dapat dimanfaatkan untuk mengoptimalkan pemerataan distribusi beban bimbingan dosen, mendorong diversifikasi topik riset, serta memanfaatkan data metrik mahasiswa sebagai sistem deteksi dini guna meningkatkan kualitas dan efisiensi penyelesaian skripsi di masa mendatang. Keterbatasan dari penelitian ini adalah penggunaan dataset tunggal yang bersifat cross-sectional dari satu program studi. Untuk penelitian selanjutnya, disarankan untuk menerapkan Temporal Network Analysis guna melihat dinamika perubahan tren riset dan interaksi dosen dari tahun ke tahun.

PUSTAKA

- Allipour Birgani, R., Takian, A., Djazayeri, A., Kianirad, A., & Pouraram, H. (2022). Climate change and food security prioritizing indices: Applying analytical hierarchy process (AHP) and social network analysis (SNA). *Sustainability*, 14(14), 8494.
- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 1–21.
- Bail, C. A. (2021). Text as data. *Sociological Methodology*, 51(1), 1–13.
- Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. Wiley

- Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1), e1230.
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of *Computers & Education*. *Computers & Education*, 151, 103855.
- Crossette, N., Carr, L. D., & Wilcox, B. R. (2023). Correlations between student connectivity and academic performance: A pandemic follow-up. *Physical Review Physics Education Research*, 19(1), 010106.
- Diana, S., Wahyudi, M. H., & Rohman, M. G. (2017). Sistem Informasi Penilaian Kinerja Pegawai pada Kantor Kecamatan Sukodadi di Kabupaten Lamongan dengan Metode AHP dan Bootstrap. *Jurnal Teknika*, 9(2).
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Hasan, R., Palaniappan, S., Razalli, A. R. A., Mahmood, S., & Sarker, A. (2021). Student academic performance prediction based on machine learning algorithms. *Education and Information Technologies*, 26, 3571–3597.
- Javed, M. A., Younis, M. S., Latif, S., Qadir, J., & Baig, A. (2021). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108, 87–111.
- Kurniawan, D., Iriani, A., & Manongga, D. (2020). Pemanfaatan Social Network Analysis (SNA) untuk menganalisis kolaborasi karyawan pada PT. Arum Mandiri Group. *Jurnal Transformatika*, 17(2), 149–158.
- Luan, H., & Tsai, C. C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society*, 24(1), 250–266.
- Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237.
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020.
- Pavlopoulos, G. A., Balauras, P., Ruiz, S. N., & Iliopoulos, I. (2017). Empirical comparison of visualization tools for larger-scale network analysis. *Advances in Bioinformatics*, 2017, 1–14.
- Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(4), 1042.
- Saqr, M., & López-Pernas, S. (2021). The temporal dynamics of online social networks: Why do students stay connected?. *Computers & Education*, 172, 104271.
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160.
- Sembiring, I., Suharyadi, S., Iriani, A., Ginting, J. V., & Ginting, J. A. (2023). A novel approach to network forensic analysis: Combining packet capture data and social network analysis. *International Journal of Advanced Computer Science and Applications*, 14(3), 234–241.
- Setiawan, E. I., Susanto, C. P., Santoso, J., Sumpeno, S., & Purnomo, M. H. (2016). Preliminary study of spam profile detection for social media using Markov Clustering: Case study on Javanese people. *2016 International Computer Science and Engineering Conference (ICSEC)*, 1–4.
- Setiawan, M., Gunawan, G., & Ferdinandus, F. X. (2021). Disjoint community detection pada network kegiatan kemahasiswaan di ISTTS menggunakan fast greedy dan walktrap. *Journal of Intelligent System and Computation*, 3(1), 7–15.
- Tuhuteru, H., & Iriani, A. (2018). Analisis kolaborasi penelitian ilmiah dosen Fakultas X dengan Social Network Analysis (SNA). *Jurnal Teknik Informatika dan Sistem Informasi*, 4(1), 31–40.
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 1–19.
- Zhang, X., Pan, Y., Wang, Y., Xu, C., & Sun, Y. (2023). Online news media analysis on information management of 'G20 Summit' based on social network analysis. *Systems*, 11(6), 290.
- Zhao, L., & Li, C. (2023). A systematic review of text mining in education. *Education and Information Technologies*, 1–35.