TESIS

ANALISIS LAPORAN BEBAN KERJA DOSEN PADA BIDANG PENELITIAN DAN PKM DENGAN MENGGUNAKAN METODE *CLUSTERING*

Disusun oleh:

Nama	**: Muhammad Egy Pratama**
NIM	**: 23.55.1449**
Konsentrasi	**: Business Intelligence**

PROGRAM STUDI S2 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

TESIS

ANALISIS LAPORAN BEBAN KERJA DOSEN PADA BIDANG PENELITIAN DAN PKM DENGAN MENGGUNAKAN METODE *CLUSTERING*

THE ANALYSIS OF LECTURER WORKLOAD IN THE FIELD OF RESEARCH AND COMMUNITY SERVICE BY USING CLUSTERING METHOD

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister

Disusun oleh:

Nama	**: Muhammad Egy Pratama**
NIM	**: 23.55.1449**
Konsentrasi	**: Business Intelligence**

PROGRAM STUDI S2 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

ANALISIS LAPORAN BEBAN KERJA DOSEN PADA BIDANG PENELITIAN DAN PKM DENGAN MENGGUNAKAN METODE *CLUSTERING*

THE ANALYSIS OF LECTURER WORKLOAD IN THE FIELD OF RESEARCH AND COMMUNITY SERVICE BY USING CLUSTERING METHOD

Dipersiapkan dan Disusun oleh

Muhammad Egy Pratama

23.55.1449

Telah disetujui oleh Dosen Pembimbing Tesis
pada hari Kamis, 07 Mei 2026

Dosen Pembimbing,

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PENGESAHAAN

ANALISIS LAPORAN BEBAN KERJA DOSEN PADA BIDANG PENELITIAN DAN PKM DENGAN MENGGUNAKAN METODE *CLUSTERING*

THE ANALYSIS OF LECTURER WORKLOAD IN THE FIELD OF RESEARCH AND COMMUNITY SERVICE BY USING CLUSTERING METHOD

Dipersiapkan dan Disusun oleh

Muhammad Egy Pratama

23.55.1449

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 07 Mei 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Dr. Andi Sunyoto, M.Kom. **NIK. 190302052**	
Dr. Kumara Ari Yuana, S.T., M.T. **NIK. 190302575**	
Prof. Dr. Kusrini, M.Kom. **NIK. 190302106**	

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 07 Mei 2026
Dekan Fakultas Ilmu Komputer

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Egy Pratama
NIM : 23.55.1449
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Analisis Laporan Beban Kerja Dosen pada Bidang Penelitian Dan PKM dengan Menggunakan Metode Clustering

Dosen Pembimbing Utama : Prof. Dr. Kusrini, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 07 Mei 2026
Yang Menyatakan,

Muhammad Egy Pratama

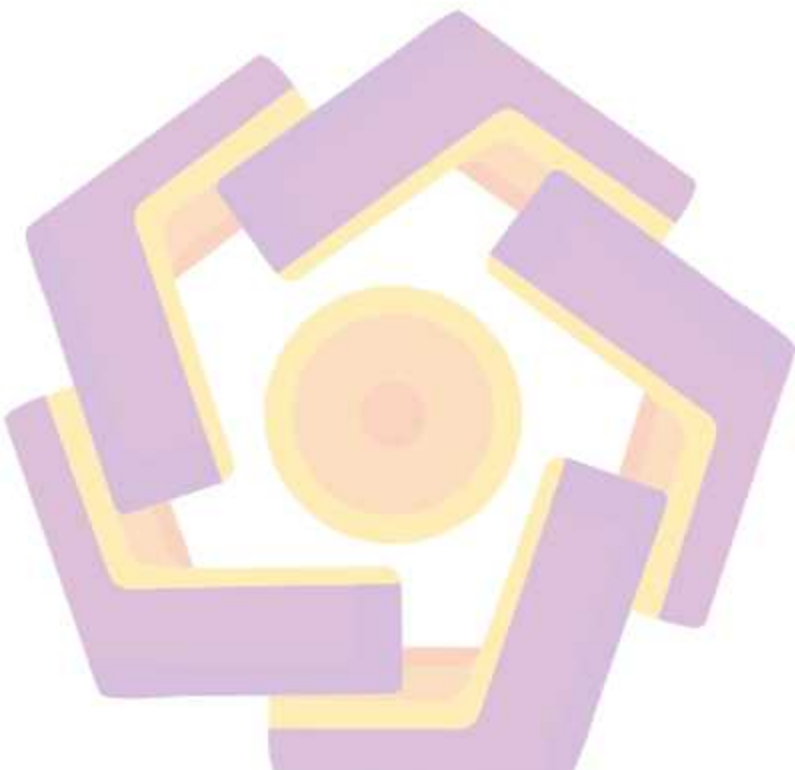
HALAMAN PERSEMBAHAN

Puji Syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah, sehingga penulis masih diberikan kesempatan untuk menyelesaikan Tesis ini, sebagai salah satu syarat untuk mendapatkan gelar Megister. Meskipun jauh dari kata sempurna, namun penulis bangga telah mencapai pada titik ini, yang akhirnya Tesis ini bisa selesai diwaktu yang tepat.

Tesis ini saya persembahkan untuk beberapa nama-nama berikut ini :

1. Ayahanda Alm. Asmaril dan Ibunda Hj. Nurhalah, S.Pd., Gr. terimakasih atas doa, semangat, motivasi, pengorbanan, nasehat serta kasih sayang yang tidak pernah henti sampai saat ini.
2. Istri Anis Komariah, M.Pd. yang selalu memberikan dukungan dalam ketidak semangatan menulis penelitian menjadi bersamangat untuk menyelesaikan kembali.
3. Kakek H. Iskandar SAG yang turut mendoakan cucunya menyelesaikan pendidikan megister ini harapanya segera menyelesaikan Tesis ini.
4. Mertua Ayahanda Slamet, S.M dan Ibunda Ngatripah yang selalu mendoakan anaknya agar dapat menyelesaikan penelitian Tesis ini.
5. Keluarga Besar Ibunda dan semua keluarga yang tidak bisa disebutkan satu-persatu, terimakasih untuk doa, nasehat, masukan dan semangatnya selama ini.
6. Pembimbing Tesis Prof. Dr. Kusrini, M.Kom. yang pembimbing tanpa lelah dan pamrih serta selalu memberikan feedback dalam Tesis ini.
7. Kucingku Kuro, Martin dan Pipip yang menemaniku ketika lagi tidak bisa melanjutkan tulisanku.

8. Semua teman-teman PJJ Informatika 23 Busines Intelegent
9. Kepada semua teman-teman, saudara yang tidak bisa saya sebutkan satu persatu, saya persembahkan Tesis ini untuk kalian semua.



HALAMAN MOTTO

"Man jadda wajada" (مَنْ جَدَّ وَجَدَ)

"Siapa yang bersungguh-sungguh, maka ia akan berhasil"

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga tesis yang berjudul **"Analisis Laporan Beban Kerja Dosen pada Bidang Penelitian dan PKM dengan Menggunakan Metode Clustering"** dapat diselesaikan dengan baik sebagai salah satu syarat untuk menyelesaikan studi pada jenjang Magister di Universitas AMIKOM Yogyakarta.

Penyusunan tesis ini tidak terlepas dari bantuan, bimbingan, serta dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Direktur Program Pascasarjana, Ketua Program Studi S2 PJJ Informatika Universitas AMIKOM Yogyakarta serta sebagai Pembimbing I yang telah memberikan arahan, bimbingan, serta masukan selama proses penyusunan tesis.
3. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom., selaku Wakil Direktur Universitas AMIKOM.
4. Bapak I Made Artha Agastya, M.Eng., Ph.D., selaku Sekretaris Program Studi S2 PJJ Informatika Dosen Pembimbing II yang telah memberikan arahan, bimbingan, serta masukan selama proses penyusunan tesis.

5. Ibu Tutut Heryanti, S.Kom, selaku Staff Pengelola Pascasarjana Administrasi dan Keuangan yang telah meluangkan waktu untuk memberikan arahan dalam pembayaran dan arahan proses administrasi.
6. Orang tua, Istri dan keluarga penulis yang senantiasa memberikan doa, dukungan moral, serta motivasi selama menempuh pendidikan.
7. Kepala LPM UINSI Samarinda Dr. Nur Kholik Affandi, M.Pd., dan seluruh Jajaran telah menginjinkan saya untuk memberikan data penelitian
8. Seluruh pihak yang terkait dan telah memberikan bantuan serta dukungan dalam penyelesaian tesis ini.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan di masa mendatang. Semoga tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan pihak-pihak yang berkepentingan.

Balikpapan, 18 Januari 2026

Muhammad Egy Pratama

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Penelitian ini menganalisis Laporan Beban Kerja Dosen (BKD) pada bidang Penelitian dan Pengabdian kepada Masyarakat (PkM) di UIN Sultan Aji Muhammad Idris Samarinda menggunakan metode K-Means clustering untuk mendukung evaluasi kinerja dan akreditasi BAN-PT APT. Data yang digunakan mencakup 2.762 aktivitas dari 214 dosen selama 6 semester (2021–2024). Hasil analisis menunjukkan *clustering* optimal dengan K=2 dan Silhouette Score 0,3882 (kategori cukup baik). Teridentifikasi dua klaster produktivitas: klaster produktivitas tinggi (54 dosen, 25,2%) dengan rata-rata total SKS 27,07, dan klaster produktivitas sedang/rendah (160 dosen, 74,8%) dengan rata-rata total SKS 12,37. Uji beda Mann-Whitney menunjukkan perbedaan signifikan antar klaster (p-value = 0,000). Evaluasi capaian KPI/IKU menunjukkan bahwa empat dari lima indikator tercapai, namun indikator publikasi internasional (80,8%) dan internasional bereputasi tinggi (88,5%) belum memenuhi target. Proyeksi skor akreditasi BAN-PT APT diperoleh 84,6 dari 400 dengan predikat C (Baik). Penelitian juga mengungkap data quality issue sebesar 81,5% aktivitas tidak terklasifikasi secara eksplisit, sehingga diperlukan standardisasi pelaporan BKD. K-Means clustering terbukti efektif sebagai dasar pengambilan keputusan strategis dalam peningkatan produktivitas dosen meskipun terbatas oleh kualitas data.

Kata kunci : Beban Kerja Dosen, K-Means Clustering, Penelitian, Pengabdian Kepada Masyarakat

ABSTRACT

This study analyzes the Lecturer Workload Report (BKD) in the Research and Community Service (PkM) field at UIN Sultan Aji Muhammad Idris Samarinda using the K-Means clustering method to support performance evaluation and BAN-PT APT accreditation. The data used includes 2,762 activities from 214 lecturers over 6 semesters (2021–2024). The analysis results show optimal clustering with K=2 and a Silhouette Score of 0.3882 (fair clustering quality). Two productivity clusters were identified: a high productivity cluster (54 lecturers, 25.2%) with an average total credits of 27.07, and a medium/low productivity cluster (160 lecturers, 74.8%) with an average total credits of 12.37. The Mann-Whitney test showed a significant difference between clusters (p-value = 0.000). Evaluation of KPI achievement indicates that four out of five indicators were met, but international publication (80.8%) and high-reputation international publication (88.5%) targets were not achieved. The projected BAN-PT APT accreditation score is 84.6 out of 400, corresponding to a C (Fair) predicate. The study also reveals a significant data quality issue: 81.5% of activities could not be explicitly classified, necessitating standardization of BKD reporting. K-Means clustering proves effective as a basis for strategic decision-making in improving lecturer productivity, despite data quality limitations.

Keywords : Lecturer Workload, K-Means Clustering, Research, Community Service

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi di berbagai sektor memiliki dampak signifikan terhadap kemajuan dan perkembangan suatu organisasi. Kemajuan dalam konteks ini mencakup perubahan paradigma dalam hal efisiensi serta peningkatan efektivitas pengelolaan waktu dan sumber daya manusia [1]. Peran teknologi menjadi salah satu faktor pendukung dalam membantu tugas Perguruan Tinggi untuk memonitoring dan mengevaluasi aktivitas Dosen dan Tenaga kependidikan.

Berdasarkan Undang-undang nomor 14 Tahun 2005 tentang Guru dan Dosen (UU Guru dan Dosen), yang mana dosen merupakan pendidik profesional dan ilmuwan dengan tugas utama mentransformasikan, mengembangkan, dan menyebarluaskan ilmu pengetahuan, teknologi, dan seni melalui pendidikan, penelitian dan pengabdian masyarakat [2]. Dosen memiliki peran yang sangat strategis dalam menentukan mutu pendidikan yang memerlukan persyaratan kemampuan profesional yang dapat dipertanggungjawabkan.

Dosen setiap semester wajib mengumpulkan Beban Kerja Dosen (BKD). BKD adalah laporan beban kerja dosen yang mencakup komponen pelaksanaan Pendidikan, Penelitian dan Pengabdian kepada Masyarakat serta unsur Penunjang kegiatan tridharma, dan atau tugas tambahan dalam kurun waktu tertentu. BKD wajib dilaporkan dengan ketentuan paling sedikit 12 SKS dan paling banyak 16 SKS [2].

Beberapa penelitian terkait dengan Sistem informasi BKD yang dirancang dengan metode UML diharapkan dapat membantu dan mempermudah proses penginputan BKD dengan menyediakan tampilan *user interface* yang jelas dan struktur data yang terorganisir. Perancangan sistem informasi BKD sangat diperlukan untuk memastikan bahwa kebutuhan pengguna terpenuhi dan proses manajemen beban kerja dosen dapat berjalan dengan efisien [3], pembuatan sistem informasi Beban Kerja Dosen (BKD) dengan metode perancangan menggunakan pemrograman berbasis web yang responsif [4]. Penelitian yang berkaitan dengan Analisis Beban Kerja Dosen dapat dilakukan secara daring dengan *Technology Acceptance Model (TAM)* untuk membantu menganalisis BKD [5], selain itu penelitian lain dilakukan dengan menggunakan Metode VORD dan penggunaan *Proto Personas* dalam menganalisis kebutuhan sistem informasi yang bertujuan untuk memfokuskan sudut pandang pengguna sistem dan juga untuk memastikan semua kebutuhan sistem terdefinisi dengan jelas untuk pengembangan perangkat lunak yang berkualitas [6]. Studi-studi ini memberikan kontribusi signifikan pada tahap pengadaan dan adopsi sistem, namun belum menyentuh tahap selanjutnya, yaitu bagaimana memanfaatkan data yang telah terkumpul dari sistem tersebut untuk menghasilkan wawasan strategis.

Di sisi lain telah berkembang pula penelitian yang berfokus pada analisis data produktivitas dosen menggunakan teknik data mining. Salah satunya adalah studi yang dilakukan oleh [7] yang mengimplementasikan K-Means clustering untuk mengelompokkan 45 dosen di Una'im Yapis Wamena berdasarkan produktivitas penelitian mereka. Penelitian tersebut berhasil mengidentifikasi tiga

klaster produktivitas (tinggi, sedang, rendah) dan memberikan rekomendasi pembinaan. Meskipun demikian, penelitian tersebut masih memiliki keterbatasan, antara lain: (1) hanya berfokus pada aspek penelitian, belum mengintegrasikan Pengabdian kepada Masyarakat (PkM) sebagai bagian dari Tridharma, (2) belum menghubungkan hasil clustering secara eksplisit dengan target kinerja institusi seperti KPI/IKU atau standar akreditasi BAN-PT APT, dan (3) penentuan jumlah cluster (K=3) dilakukan tanpa validasi objektif seperti Elbow Method atau Silhouette Score.

Penelitian ini hadir untuk mengatasi keterbatasan tersebut dengan melakukan analisis clustering pada data BKD bidang penelitian dan PkM di UINSI Samarinda yang terdiri dari 214 dosen dengan 2.762 records aktivitas selama 6 semester (2021-2024). Jumlah cluster optimal ditentukan secara objektif menggunakan Elbow Method dan Silhouette Score.

Penelitian ini menggunakan algoritma **K-Means clustering** karena data BKD yang telah dinormalisasi memenuhi asumsi *spherical* (struktur data cenderung *globular*). Hasil analisis Principal Component Analysis (PCA) menunjukkan bahwa dua komponen utama mampu menjelaskan **82,1% varians data**, mengindikasikan bahwa data memiliki bentuk yang sesuai untuk K-Means. Selain itu, K-Means dipilih karena efisiensinya dalam menangani data berukuran besar (2.762 aktivitas) dan kemudahan interpretasi hasilnya bagi pengambil kebijakan di perguruan tinggi.

Namun, penelitian ini juga mengidentifikasi tantangan data quality yang signifikan: 81.5% dari aktivitas yang tercatat tidak memiliki klasifikasi tipe

publikasi yang eksplisit (jurnal, prosiding, book chapter, atau sejenisnya). Hal ini menunjukkan perlunya peningkatan standar data governance dalam sistem E-BKD. Meskipun demikian, hasil clustering tetap valid secara metodologis dan dapat digunakan sebagai "decision aid" untuk strategi pengembangan dosen, terutama ketika dikombinasikan dengan qualitative assessment dari faculty leadership. Hasil clustering selanjutnya digunakan untuk mengevaluasi capaian KPI/IKU institusi serta memproyeksikan pemenuhan kriteria akreditasi BAN-PT APT.

Kondisi serupa dialami oleh Universitas Islam Negeri Sultan Aji Muhammad Idris (UINSI) Samarinda, yang mana Lembaga Penjaminan Mutu (LPM) telah mengembangkan aplikasi E-BKD untuk pelaporan administratif Dosen, yang berhasil mengumpulkan ribuan data aktivitas dosen pada bidang Pendidikan, Penelitian, Pengabdian Kepada Masyarakat dan penunjang. Namun pemanfaatan data ini masih terbatas pada fungsi administratif, seperti verifikasi pemenuhan beban minimal dan pengarsipan. Akibatnya manajemen Universitas menghadapi tantangan yang pertama memetakan kinerja dosen secara objektif siapa yang telah berkontribusi signifikan terhadap publikasi berkualitas, dan siapa yang membutuhkan pendampingan, kedua mengukur pencapaian target strategis untuk mengevaluasi secara berkala sejauh mana capaian IKU/KPI institusi, khususnya bidang publikasi ilmiah, yang menjadi komponen penilaian akreditasi BAN-PT APT, ketiga merumuskan kebijakan yang tepat sasaran program peningkatan kapasitas dosen cenderung bersifat umum dan kurang efektif karena tidak membedakan kebutuhan kelompok dosen yang berbeda.

Untuk mengatasi permasalahan tersebut, diperlukan suatu pendekatan analisis yang mampu mengolah data BKD yang besar dan kompleks menjadi informasi yang terstruktur dan bermakna. Pendekatan yang tepat adalah menggunakan data mining, khususnya teknik clustering. Clustering adalah metode pengelompokan data berdasarkan kesamaan karakteristik, yang bertujuan untuk menemukan pola atau kelompok alami dalam suatu dataset [8]. Dalam hal ini clustering dapat mengelompokkan dosen berdasarkan profil produktivitas Penelitian dan PKM.

Penelitian ini memilih algoritma K-Means clustering karena beberapa alasan yang pertama kemampuan dalam menangani data dalam jumlah besar secara cepat dan efisien [9], kedua kesederhanaan dan kemudahan interpretasi hasilnya [10] ketiga serta telah terbukti efektif dalam berbagai aplikasi pengelompokan di bidang pendidikan, seperti pengelompokan siswa berprestasi [10] atau pemetaan produktivitas akademik. Dengan menerapkan K-Means pada data BKD UIN Sultan Aji Muhammad Idris Samarinda, penelitian ini diharapkan dapat menghasilkan pengelompokan dosen ke dalam dua kategori produktivitas yang berbeda secara signifikan (produktivitas tinggi dan produktivitas sedang/rendah) sebagai dasar penyusunan strategi pembinaan yang tepat sasaran.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka rumusan masalah pada penelitian ini adalah:

1) Bagaimana metode *clustering* dapat digunakan untuk menganalisis Laporan Beban Kerja Dosen pada bidang penelitian dan PkM dalam pemenuhan kriteria

akreditasi BAN-PT APT di Universitas Islam Sultan Aji Muhammad Idris Samarinda?

2) Bagaimana hasil analisis K-Means *clustering* pada Laporan Beban Kerja Dosen pada bidang penelitian dan PkM dapat menentukan capaian *Key Performance Indicators* (KPI) atau Indikator Kinerja Utama (IKU) di Universitas Islam Sultan Aji Muhammad Idris Samarinda?

1.3. Batasan Masalah

Batasan Masalah dalam penelitian ini adalah:

1) Mengunakan teknologi *clustering*, khususnya metode K-Means, dapat diterapkan untuk menganalisis dan mengelompokkan data BKD guna menemukan pola atau tren dalam beban kerja dosen.
2) Penggunaan evaluasi BKD dengan metode K-Means dapat memudahkan pemantauan kinerja dosen serta membantu dalam evaluasi pemenuhan kriteria akreditasi.
3) Data yang digunakan dalam penelitian ini berdasarkan pada data BKD dosen LPM Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda.
4) Evaluasi kinerja sistem akan dilakukan berdasarkan ketersediaan data dan *feedback* dari pengguna, dengan fokus pada efisiensi, kehandalan, dan kegunaan aplikasi.
5) Penelitian tidak akan membahas secara mendalam aspek hukum atau administratif terkait pelaporan BKD, namun akan berfokus pada aspek

teknologi yang dapat meningkatkan manajemen dan efisiensi proses tersebut.

6) Indikator penilaian pada Penelitian ini berdasarkan Matrik BAN-PT Akreditasi Perguruan Tinggi.
7) Key Performance Indicators (KPI) atau Indikator Kinerja Utama (IKU) UIN Sultan Aji Muhammad Idris Samarida
8) Penelitian ini akan mengambil dataset yang hanya berfokus pada semester ganjil 2021/2022 sampai semester genap 2023/2024 selama 6 semester.
9) Penilai ini berfokus pada pengelompokan Penelitian dan PkM yang sesuai dengan Matrik BAN-PT Akreditasi Perguruan Tinggi dan Key Performance Indicators (KPI) atau Indikator Kinerja Utama (IKU) UIN Sultan Aji Muhammad Idris Samarida

1.4. Tujuan Penelitian

1) Bagaimana metode *clustering* dapat digunakan untuk menganalisis Laporan Beban Kerja Dosen pada bidang penelitian dan PkM dalam pemenuhan kriteria akreditasi BAN-PT APT di Universitas Islam Sultan Aji Muhammad Idris Samarinda.
2) Bagaimana hasil analisis K-Means *clustering* pada Laporan Beban Kerja Dosen pada bidang penelitian dan PkM dapat menentukan capaian Key Performance Indicators (KPI) atau Indikator Kinerja Utama (IKU) di Universitas Islam Sultan Aji Muhammad Idris Samarinda.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1) Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan teori dan pemahaman dalam penggunaan metode *clustering*, khususnya algoritma K-Means, dalam konteks pengelolaan dan analisis data di bidang penelitian dan PkM. Secara spesifik, penelitian ini memperkaya khazanah ilmu pengetahuan tentang penerapan *data mining* di perguruan tinggi, terutama dalam hal menghubungkan hasil *clustering* dengan indikator kinerja strategis seperti KPI/IKU dan standar akreditasi BAN-PT, yang masih jarang dilakukan dalam penelitian-penelitian sebelumnya. Selain itu, penelitian ini juga dapat menambah wawasan tentang bagaimana teknik *machine learning* dapat diterapkan secara efektif sebagai dasar dalam pengambilan keputusan strategis di perguruan tinggi.
2) Penelitian ini memberikan manfaat praktis bagi Universitas Islam Negeri Sultan Aji Muhammad Idris Samarinda dengan:
   a. Menghasilkan pemetaan (klasterisasi) dosen berdasarkan tingkat produktivitas penelitian dan PkM, sehingga institusi dapat mengidentifikasi dosen dengan produktivitas tinggi, sedang, dan perlu pembinaan secara objektif.
   b. Menyediakan evaluasi capaian KPI/IKU berbasis data, yang selama ini sulit dilakukan karena keterbatasan alat analisis.
   c. Memberikan proyeksi pemenuhan kriteria akreditasi BAN-PT APT, yang dapat menjadi acuan dalam perencanaan strategis dan peningkatan kinerja institusi.

Dengan demikian, penelitian ini tidak hanya membantu dalam merumuskan kebijakan yang lebih tepat sasaran, tetapi juga mendukung upaya peningkatan mutu dan daya saing institusi di tingkat nasional.

1.6. Hipotesis

Berdasarkan rumusan masalah, tujuan penelitian, dan tinjauan pustaka yang telah diuraikan, hipotesis dalam penelitian ini adalah sebagai berikut:

1. Metode K-Means clustering dapat diterapkan secara efektif untuk menganalisis dan mengelompokkan data laporan Beban Kerja Dosen (BKD) pada bidang penelitian dan Pengabdian kepada Masyarakat (PkM) di UIN Sultan Aji Muhammad Idris Samarinda dengan menghasilkan kualitas clustering yang cukup baik (*Silhouette Score* = 0,3882, kategori 0,3–0,5), sehingga dapat mendukung pemenuhan kriteria akreditasi BAN-PT APT.
2. Terdapat perbedaan signifikan dalam pola produktivitas penelitian dan PkM antar kelompok dosen yang dihasilkan dari proses *clustering*, di mana kelompok dengan produktivitas tinggi memiliki kontribusi publikasi yang lebih tinggi dibandingkan kelompok produktivitas sedang, khususnya dalam kategori publikasi berkualitas (internasional bereputasi dan nasional terakreditasi).

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian tentang Beban Kerja Dosen (BKD) telah banyak dilakukan, terutama dalam konteks pengembangan sistem informasi. [3] merancang sistem informasi BKD berbasis web menggunakan metode UML untuk memudahkan proses penginputan dan pelaporan. [4]) mengimplementasikan sistem serupa dengan pendekatan pemrograman web responsif di Politeknik Negeri Bengkalis. Kedua penelitian tersebut berfokus pada aspek teknis dan efisiensi administrasi, namun belum menyentuh analisis data yang dihasilkan. Sementara itu, [5] menggunakan Technology Acceptance Model (TAM) untuk menganalisis penerimaan dosen terhadap aplikasi BKD online, dan [6] melakukan analisis kebutuhan sistem dengan metode VORD dan Proto Personas. Studi-studi ini berkontribusi pada tahap pengadaan dan adopsi sistem, namun belum memanfaatkan data yang terkumpul untuk menghasilkan wawasan strategis.

Di sisi lain, algoritma K-Means clustering telah banyak diaplikasikan di bidang pendidikan. [10] menggunakan K-Means untuk mengelompokkan siswa berdasarkan nilai mata pelajaran dan absensi, kemudian mengidentifikasi siswa unggul dengan metode Simple Additive Weighting (SAW). Penelitian ini membuktikan efektivitas K-Means dalam mengelompokkan entitas pendidikan berdasarkan karakteristik kinerja. [11] juga menerapkan K-Means untuk

mengelompokkan penilaian dosen berdasarkan indeks kepuasan mahasiswa, yang menghasilkan dua cluster (baik dan kurang) serta rekomendasi peningkatan kinerja dosen. [12] menggunakan algoritma yang sama untuk mengelompokkan calon penerima beasiswa, menunjukkan fleksibilitas K-Means dalam berbagai konteks pengambilan keputusan.

Dalam konteks yang lebih spesifik, [13] menerapkan K-Means untuk mengelompokkan produktivitas tanaman padi di Jawa Tengah. Meskipun domainnya berbeda, penelitian ini menunjukkan bahwa K-Means efektif untuk analisis produktivitas berbasis wilayah.

Penelitian yang paling relevan dengan kajian ini adalah studi yang dilakukan oleh [7] yang mengimplementasikan K-Means clustering untuk mengelompokkan produktivitas penelitian 45 dosen di Una'im Yapis Wamena. Menggunakan variabel publikasi, sitasi, h-index, penelitian didanai, dan kolaborasi, penelitian tersebut berhasil mengidentifikasi tiga klaster produktivitas (tinggi, sedang, rendah). Namun, penelitian ini memiliki beberapa keterbatasan: (1) hanya berfokus pada aspek penelitian, belum mengintegrasikan Pengabdian kepada Masyarakat (PkM); (2) belum menghubungkan hasil clustering dengan target kinerja institusi seperti KPI/IKU atau standar akreditasi; dan (3) penentuan jumlah cluster (K=3) dilakukan tanpa validasi objektif seperti Elbow Method atau Silhouette Score.

2.2. Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian Analisis Laporan Beban Kerja Dosen pada Bidang Penelitian dan PKM dengan Menggunakan Metode Clustering

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Implementasi K-Means Clustering untuk Mengelompokkan Data Produktivitas Penelitian Dosen di Una'im Yapis Wamena	1.Syarifah 2.Muhammad Ali 3.Asman Samad 4.Fayra Nabillah Azza 5.Hasriani M 6.Susiana Media: JBIDAI, 2025 [7]	Mengelompokkan produktivitas penelitian dosen ke dalam kategori tinggi, sedang, dan rendah untuk membantu institusi dalam menentukan kebijakan pembinaan yang tepat sasaran.	Berhasil mengelompokkan 45 dosen menjadi 3 cluster: 1. Cluster 1 (Tinggi): 8 dosen (17,8%) dengan rata-rata 15-18 publikasi 2. Cluster 2 (Sedang): 25 dosen (55,6%) dengan rata-rata 4-7 publikasi 3. Cluster 3 (Rendah): 12 dosen (26,6%) dengan rata-rata 0-2 publikasi 4. Silhouette Score: 0,487 (kualitas cukup baik)	**Kelemahan:** 1. Hanya berfokus pada aspek penelitian, belum mengintegrasikan PkM 2. Belum menghubungkan hasil clustering dengan target kinerja institusi (KPI/IKU) atau standar akreditasi 3. Penentuan jumlah cluster (K=3) dilakukan tanpa validasi objektif seperti Elbow Method atau Silhouette Score	Penelitian [7] 1. Fokus: Produktivitas penelitian dosen 2. Metode: K-Means clustering dengan K=3 (tanpa validasi objektif) 3. Data: 45 dosen, 5 variabel (publikasi, sitasi, h-index, penelitian didanai, kolaborasi) Sedangkan Penelitian pada Tesis Ini: 1. Fokus: Produktivitas penelitian DAN PkM dosen 2. Metode : *K-Means clustering* dengan validasi objektif (*Elbow + Silhouette*), menghasilkan K=2 optimal dengan skor 0,3882 (kategori cukup baik). Meskipun nilai ini lebih rendah dari penelitian Ali dkk. (0,487), penelitian ini menggunakan jumlah sampel lebih besar (214 vs 45 dosen) dan cakupan lebih luas

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					(meskipun nilai akhir dilaporkan) 4. Jumlah sampel relatif kecil (45 dosen) **Saran:** Penelitian selanjutnya disarankan untuk memperluas cakupan pada bidang lainnya dan mengaitkan dengan indikator kinerja institusi.	(penelitian + PKM). 3. Data: 214 dosen dengan 2.762 records (lebih besar) 4. Hasil clustering digunakan untuk mengevaluasi KPI/IKU dan memproyeksikan akreditasi BAN-PT APT Perbedaan Utama: Perlu ditekankan bahwa perbandingan nilai metrik (misalnya Silhouette Score) antara penelitian ini dan penelitian Ali et al. (2025) tidak bersifat *apple-to-apple* karena perbedaan jumlah sampel (214 vs 45 dosen), cakupan (penelitian dan PKM vs hanya penelitian), periode data (6 semester vs tidak disebutkan), serta karakteristik institusi. Oleh karena itu, perbandingan lebih difokuskan pada pola temuan (adanya gap produktivitas, konsentrasi kinerja pada sebagian kecil dosen) daripada perbandingan nilai absolut metrik.
2	Pengembangan Sistem Informasi Pemantauan	1. Pahrul Irfan,	Untuk mengembangkan dan menerapkan sistem informasi	Kegiatan pengabdian berhasil mencapai tujuannya dengan menghasilkan sistem	Saran: 1. Melakukan pemantauan berkala dan	Penelitian [14] 1. Fokus: Pengembangan sistem informasi pemantauan beban kerja

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Beban Kerja Dosen Terintegrasi Berbasis Prototipe *(Development of an Integrated Lecturer Workload Monitoring Information System Using the Prototype Model)*	2. I Gede Pasek Suta Wijaya, 3. Halil Akhyar, 4. Ariyan Zubaidi, 5. Ahmad Zafrullah Media Publikasi Jurnal: JBegaTI 2025 [14]	pemantauan beban kerja dosen yang terintegrasi guna mengatasi permasalahan pencatatan manual di Program Studi Teknik Informatika, Universitas Mataram. Tujuannya adalah meningkatkan efisiensi, akurasi data, transparansi distribusi tugas, serta mendukung evaluasi dan pengambilan keputusan berbasis data di tingkat program studi.	informasi pemantauan beban kerja dosen yang fungsional. Pengujian *Blackbox* memastikan seluruh fitur utama berjalan sesuai spesifikasi. Uji *System Usability Scale* (SUS) terhadap delapan responden menghasilkan skor rata-rata 84,06 yang masuk kategori Excellent, membuktikan bahwa sistem mudah digunakan dan diterima dengan baik oleh pengguna. Integrasi pendekatan partisipatif dan metode prototipe terbukti efektif dalam menciptakan solusi yang tepat guna dan berkelanjutan.	pelatihan penyegaran bagi pengguna. 2. Mengembangkan fitur tambahan sesuai kebutuhan di masa mendatang. 3. Menerapkan sistem secara lebih luas di program studi lain di lingkungan fakultas maupun universitas. 4. Mengintegrasikan sistem dengan Sistem SKST (Sumber Karya Statuta?) agar penugasan dosen di luar Tridarma (seperti tugas struktural) turut tercatat secara otomatis, sehingga memberikan gambaran beban kerja yang lebih	dosen yang terintegrasi berbasis prototipe. 2. Metode: Prototipe dengan pendekatan partisipatif dan pengujian *Blackbox* serta *System Usability Scale* (SUS). 3. Tujuan: Menyediakan alat untuk pencatatan dan pemantauan BKD yang efisien, akurat, dan transparan di tingkat program studi. Sedangkan Penelitian pada Tesis Ini: 1. Fokus: Analisis data historis BKD untuk mengevaluasi produktivitas dosen dalam bidang penelitian dan PkM. 2. Metode: *Data mining* dengan algoritma K-Means *clustering* serta validasi menggunakan *Elbow Method* dan *Silhouette Score*. 3. Perbedaan Utama: Penelitian Irfan dkk. berfokus pada penyediaan infrastruktur teknologi (sistem) untuk mencatat dan memantau BKD. Penelitian ini berfokus pada ekstraksi pengetahuan dari data BKD yang telah terkumpul

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					komprehensif. Kelemahan: 1. Pengujian SUS hanya melibatkan 8 responden, sehingga generalisasi tingkat penerimaan pengguna mungkin terbatas. 2. Implementasi awal baru terbatas pada satu program studi (Teknik Informatika, Universitas Mataram). 3. Keberlanjutan sistem sangat bergantung pada komitmen mitra dan sumber daya yang tersedia pasca-kegiatan pengabdian.	untuk evaluasi kinerja dosen, pengukuran KPI, dan penilaian akreditasi institusi. Dengan kata lain, penelitian Irfan dkk. membangun alat untuk menghasilkan data, sedangkan penelitian ini memanfaatkan data tersebut untuk pengambilan keputusan strategis di tingkat institusi.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	A K-means Clustering Analysis to Assess Educators' and Students' Perceptions of Digital Technology in Higher Education *(Analisis Klastering K-Means untuk Menilai Persepsi Pendidik dan Siswa terhadap Teknologi Digital di Pendidikan Tinggi)*	1. Umar Faruk Abdulhamid, 2. Muhammad Aminu Ahmad, 3. Mohammed Ibrahim Media Publikasi Jurnal : A periodical of the Faculty of Natural and Applied Sciences, UMYU, Katsina, 2025	1. Menganalisis persepsi pendidik (dosen) dan mahasiswa terhadap penggunaan teknologi digital dalam pengajaran dan pembelajaran di institusi pendidikan tinggi Nigeria. 2. Mengungkap pola perspektif pendidik dan mahasiswa tentang penggunaan teknologi digital. 3. Mengevaluasi kegunaan algoritma *K-means clustering* dalam pengenalan pola, dengan fokus pada persepsi dosen dan	Analisis K-Means berhasil mengidentifikasi tiga klaster yang berbeda untuk kedua kelompok responden (dosen dan mahasiswa): 1. Laggards (Kelompok Tertinggal): Memiliki tingkat integrasi teknologi digital yang sangat rendah dan persepsi manfaat yang sangat rendah. 2. Adorers (Kelompok Pengagum): Memiliki tingkat integrasi teknologi yang rendah tetapi memiliki persepsi manfaat yang tinggi. 3. Adopters (Kelompok Pengadopsi): Memiliki tingkat integrasi teknologi yang tinggi dan persepsi manfaat yang moderat. Mayoritas pendidik (62%) dan mahasiswa memahami manfaat	Saran untuk Penelitian Mendatang: 1. Menggunakan dataset berskala lebih besar karena algoritma *machine learning* membutuhkan volume data yang besar untuk bekerja secara efektif. 2. Menyelidiki fungsi dari berbagai algoritma klastering lainnya untuk memahami seberapa baik kinerja mereka dalam menilai persepsi dan penggunaan teknologi digital. 3. Seiring kemajuan AI dalam pendidikan, penelitian selanjutnya dapat menyelidiki	Penelitian [15] 1. Fokus: Menganalisis persepsi pendidik (dosen) dan mahasiswa terhadap penggunaan teknologi digital di pendidikan tinggi Nigeria. 2. Metode: K-Means *clustering* berdasarkan data kuesioner (data persepsi/subjektif). 3. Tujuan: Mengidentifikasi pola persepsi dan tingkat integrasi teknologi digital pada dosen dan mahasiswa (menghasilkan klaster *laggards*, *adorers*, dan *adopters*). Sedangkan Penelitian Pada Tesis Ini: 1. Fokus: Menganalisis produktivitas dosen dalam bidang penelitian dan PkM berdasarkan data BKD aktual. 2. Metode: K-Means *clustering* berdasarkan data kinerja riil dari sistem e-BKD (data objektif/terukur). 3. Perbedaan Utama: Abdulhamid dkk. menggunakan data persepsi (kuesioner) yang bersifat subjektif, sedangkan penelitian ini menggunakan data kinerja

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			mahasiswa terhadap integrasi teknologi digital.	teknologi digital, namun masih ada kelompok signifikan yang memiliki persepsi rendah. Hal ini menyoroti perlunya sumber daya teknologi yang memadai dan strategi dukungan, termasuk pelatihan pengembangan profesional bagi pendidik.	bagaimana pandangan pemangku kepentingan terhadap penggunaan dan penerimaan AI dalam pengajaran dan pembelajaran. Kelemahan Penelitian (Implisit): 1. Penelitian ini hanya menggunakan algoritma K-Means. Perbandingan dengan algoritma klastering lain tidak dilakukan. 2. Penelitian terbatas pada institusi pendidikan tinggi di Nigeria, sehingga generalisasi temuan ke konteks geografis lain perlu	aktual dari sistem yang merekam aktivitas nyata dosen. Hasil klaster Abdulhamid dkk. menggambarkan sikap dan kesiapan terhadap teknologi, sedangkan hasil klaster penelitian ini menggambarkan produktivitas riil yang berdampak langsung pada KPI dan akreditasi institusi. Penelitian ini juga mengaitkan hasil *clustering* dengan target strategis institusi (KPI/IKU dan standar BAN-PT), yang tidak dilakukan dalam penelitian Abdulhamid dkk.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					dilakukan dengan hati-hati. 3. Beberapa konstruk dalam pengujian reliabilitas menunjukkan nilai Cronbach's Alpha yang rendah (misalnya, *Self-efficacy* = 0,216 untuk mahasiswa; *Workload* = 0,403 untuk pendidik), yang dapat mempengaruhi konsistensi internal dari instrumen penelitian.	
4	Optimasi Penentuan Sentroid Awal pada K-Means untuk meningkatkan hasil evaluasi davies-bouldin index	1. Hendrik, 2. Kusrini, 3. Kusnawi Media Publikasi Jurnal Informatika Teknologi	Untuk mengatasi permasalahan penentuan sentroid awal secara acak pada algoritma K-Means yang dapat mempengaruhi hasil klasterisasi. Penelitian ini	1. Baik metode Elbow maupun PSO sama-sama berhasil menentukan jumlah *cluster* yang paling optimal, yaitu k=3, dengan nilai DBI yang sama yaitu 0,7376. 2. Perbedaan utama terletak pada efisiensi waktu	Saran: 1. Penelitian selanjutnya disarankan untuk menggunakan dataset yang berbeda untuk	Penelitian [16] 1. Fokus: Optimasi penentuan sentroid awal pada algoritma K-Means untuk meningkatkan hasil evaluasi *Davies-Bouldin Index* (DBI). 2. Metode: Komparasi dua metode optimasi (Elbow dan Particle

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		dan Sains (JINTEKS), 2026	bertujuan mengusulkan dan membandingkan dua metode optimasi, yaitu metode Elbow dan *Particle Swarm Optimization* (PSO), untuk meningkatkan hasil evaluasi algoritma K-Means yang diukur menggunakan *Davies-Bouldin Index* (DBI).	komputasi. Metode Elbow jauh lebih unggul dengan waktu iterasi 0,297 detik, sedangkan metode PSO membutuhkan waktu yang jauh lebih lama yaitu 779 detik. 3. Dengan demikian, metode Elbow lebih efisien dan direkomendasikan untuk optimasi K-Means, terutama untuk dataset dengan ukuran yang besar.	menguji konsistensi hasil. 2. Disarankan untuk membandingkan dengan metode optimasi lainnya (seperti Genetic Algorithm, Artificial Bee Colony, dll.) agar diperoleh perbandingan evaluasi yang lebih komprehensif. Kelemahan Penelitian (Implisit): 1. Penelitian ini hanya menggunakan satu dataset (Online Retail II dari UCI Machine Learning Repository), sehingga generalisasi hasil ke dataset lain perlu diuji lebih lanjut.	Swarm Optimization) pada dataset publik (Online Retail II). 3. Tujuan: Meningkatkan efisiensi waktu komputasi dan kualitas *clustering* K-Means. Sedangkan Penelitian Pada Tesis Ini: 1. Fokus: Aplikasi K-Means *clustering* pada dataset riil BKD untuk mendukung evaluasi kinerja dan akreditasi perguruan tinggi. 2. Metode: Implementasi K-Means dengan validasi Elbow dan Silhouette pada data BKD UIN Sultan Aji Muhammad Idris Samarinda. 3. Perbedaan Utama: Penelitian Hendrik dkk. bersifat metodologis (mencari cara terbaik menjalankan K-Means dengan membandingkan teknik optimasi). Penelitian ini bersifat aplikatif/terapan (*applied research*) dengan menggunakan K-Means untuk memecahkan masalah

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					2. Penelitian ini hanya berfokus pada perbandingan dua metode optimasi (Elbow dan PSO). 3. Nilai DBI yang sama (0,7376) untuk kedua metode menunjukkan bahwa dari sisi kualitas *cluster* tida k ada peningkatan, meskipun waktu komputasi PSO jauh lebih lambat.	manajerial nyata di institusi pendidikan. Hendrik dkk. berfokus pada peningkatan performa algoritma (waktu dan nilai DBI), sedangkan penelitian ini berfokus pada peningkatan performa institusi melalui pemahaman pola produktivitas dosen dan pencapaian target akreditasi. Dataset yang digunakan juga berbeda: Hendrik dkk. menggunakan data publik sektor ritel, sedangkan penelitian ini menggunakan data internal perguruan tinggi (BKD) yang bersifat sensitif dan strategis.
5	Perancangan Sistem Informasi Beban Kerja Dosen Berbasis Web dengan UML	1. Muhamm ad Nugraha 2. Mia Rosmeida Media Publikasi: Jurnal Algoritma Tahun: 2021	Tujuan dari penelitian yang dilakukan adalah untuk menganalisis dan merancang kebutuhan sistem guna membangun sistem informasi manajemen Beban Kerja Dosen (BKD)	Berdasarkan hasil penelitian yang dilakukan, kesimpulannya adalah perancangan sistem informasi BKD berhasil dibuat dan siap untuk diimplementasikan. Diharapkan sistem ini dapat membantu para pengembang sistem di institusi terkait untuk	Saran untuk penelitian selanjutnya adalah pengembangan perancangan sistem ini dengan menambahkan fitur-fitur seperti kemampuan untuk mengedit dokumen yang terhubung langsung ke Google Docs, fitur pratinjau	Penelitian [3] 1. Fokus: Perancangan sistem informasi BKD berbasis web. 2. Metode: Pemodelan dengan Unified Modeling Language (UML). 3. Tujuan: Menghasilkan blueprint sistem untuk mempermudah

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				mengimplementasikannya dengan mudah, sehingga proses evaluasi BKD di suatu kampus dapat menjadi lebih efisien dan mudah diakses	cetak, dan fitur lainnya yang dapat lebih mempermudah pengguna. Kelemahan dari penelitian sebelumnya adalah perancangan yang kurang lengkap karena hanya mencakup pembuatan use case diagram saja, sehingga proses lainnya belum terlihat. Selain itu, pada perancangan lain fokus pada pemeriksaan dan pelaporan BKD tanpa adanya pengarsipan data	proses pelaporan dan evaluasi BKD. Sedangkan Penelitian pada Tesis Ini : 1. Fokus: Analisis konten/data dari sistem BKD yang sudah berjalan. 2. Metode: Data mining dengan algoritma K-Means clustering. 3. Perbedaan Utama: Penelitian ini bukan merancang *sistem* baru, tetapi menganalisis *data output* dari sistem yang sudah ada. Temuan digunakan untuk evaluasi kinerja dosen, pengukuran KPI, dan penilaian akreditasi institusi yang mana merupakan tahap analitik yang melampaui fungsi pelaporan sistem itu sendiri..
6	Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada	1.Normah, 2.Siti Nurajizah, 3.Arinda Salbinda. Media publikasi	Tujuan penelitian ini adalah untuk membantu Toko Helai dalam menentukan produk penjualan yang termasuk ke	Kesimpulan dari penelitian ini adalah sebagai berikut: 1. Metode K-Means dapat diterapkan pada Toko Helai untuk menentukan penjualan baju mana	Saran: 1. Penggunaan Data yang Lebih Luas: Disarankan untuk menggunakan data penjualan yang lebih luas dan	Penelitian [17] 1. Fokus: Pengelolaan inventaris bisnis ritel (swalayan). 2. Metode: K-Means clustering untuk mengelompokkan produk berdasarkan data penjualan.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Toko Fashion Hijab Banten	adalah "Jurnal Teknik Komputer AMIK BSI," dan tahun publikasinya adalah 2021.	dalam kategori sangat laris, laris, serta kurang laris. Selain itu, penelitian ini bertujuan untuk mengimplementasikan konsep data mining menggunakan Algoritma K-Means untuk meningkatkan penjualan.	yang sangat laris, laris, dan kurang laris. 2. Penerapan metode K-Means membantu Toko Helai dalam mengembangkan strategi persediaan stok baju. 3. Hasil akhir dari penerapan metode K-Means menunjukkan bahwa terdapat 11 artikel yang sangat laris, 55 artikel yang laris, dan 34 artikel yang kurang laris.	beragam untuk mendapatkan hasil yang lebih akurat dan representatif. 2. Penerapan Metode Lain: Selain K-Means, dapat dipertimbangkan untuk menerapkan metode data mining lainnya untuk membandingkan hasil dan meningkatkan akurasi analisis. 3. Peningkatan Sistem Informasi: Toko Helai perlu meningkatkan sistem informasi akuntansi dan manajemen data agar data penjualan dapat dikelola dengan lebih baik dan dapat digunakan	3. Tujuan: Efisiensi operasional mengetahui produk laris/kurang laris untuk optimasi stok Sedangkan Penelitian pada Tesis ini : 1. Fokus: Evaluasi kinerja akademik di perguruan tinggi. 2. Metode: K-Means clustering untuk mengelompokkan dosen berdasarkan produktivitas penelitian & PKM. 3. Perbedaan Utama: Meski menggunakan algoritma yang sama (K-Means), konteks penerapannya berbeda secara fundamental. Penelitian ini mengaplikasikan *clustering* untuk a nalisis sumber daya manusia dan kebijakan akademik, bukan untuk logistik bisnis. Hasil *clustering* digunakan untuk menilai pencapaian KPI institusional dan standar akreditasi nasional (BAN-PT), schingga memiliki implikasi strategis pada

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					untuk analisis lebih lanjut. Kelemahan: 1. Ketergantungan pada Data Historis: Hasil analisis sangat bergantung pada data historis yang digunakan. Jika data tidak representatif, hasilnya mungkin tidak akurat. 2. Pemilihan Centroid Awal: Pemilihan centroid awal dalam metode K-Means dapat mempengaruhi hasil akhir, dan jika tidak dilakukan dengan hati-hati, dapat menghasilkan cluster yang kurang optimal.	level kebijakan dan penjaminan mutu pendidikan tinggi.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					3. Keterbatasan dalam Mengelompokkan: Metode K-Means memiliki keterbatasan dalam mengelompokkan data yang tidak berbentuk bulat atau memiliki distribusi yang kompleks	
7	Analisis Clustering pada Pengguna Brand HP Menggunakan Metode K-Means	1.Indah Nuryani 2.Dedi Darwis. Media publikasi adalah "Prosiding Seminar Nasional Ilmu Komputer," dan tahun	Tujuan penelitian ini adalah untuk memberikan hasil penerapan algoritma K-Means berupa data penggunaan merek HP mahasiswa di wilayah Bandarlampung yang dikelompokkan ke dalam tiga cluster,	Kesimpulan dari penelitian ini adalah bahwa algoritma K-Means dapat digunakan dengan baik untuk mengelompokkan data penggunaan merek HP mahasiswa di Bandarlampung, menghasilkan tiga cluster: sangat baik, baik, dan cukup. Hasil ini dapat membantu mahasiswa dalam menentukan pilihan HP yang tepat untuk	Saran yang diberikan dalam penelitian ini adalah agar penelitian selanjutnya dapat menggunakan berbagai algoritma selain K-Means untuk mengetahui tingkat akurasi metode yang memiliki kinerja lebih baik dibandingkan dengan K-Means. Kelemahan yang diidentifikasi dalam	Penelitian [11] 1. Fokus: Preferensi konsumen terhadap merek HP di kalangan mahasiswa. 2. Metode: K-Means clustering berdasarkan data kuesioner. 3. Tujuan: Segmentasi pasar membantu mahasiswa memilih HP dan memberi referensi bagi penjual.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		publikasinya adalah 2021	yaitu sangat baik, baik, dan cukup	menunjang proses pendidikan dan juga menjadi referensi bagi penjual HP di wilayah tersebut Selain itu, penelitian ini merekomendasikan penggunaan berbagai algoritma lain di masa depan untuk meningkatkan akurasi dan kinerja analisis	penelitian ini adalah bahwa hanya menggunakan algoritma K-Means, yang mungkin membatasi pemahaman tentang data. Oleh karena itu, eksplorasi dengan algoritma lain diharapkan dapat memberikan hasil yang lebih komprehensif dan akurat.	Sedangkan Penelitian pada Tesis ini: 6. Fokus: Kinerja dan produktivitas akademik dosen. 7. Metode: K-Means clustering berdasarkan data kinerja aktual dari sistem BKD 8. Perbedaan Utama: Penelitian ini menggunakan data kinerja objektif yang tercatat sistem, bukan data preferensi subjektif dari kuesioner. Aplikasi clustering di sini bukan untuk analisis pasar, tetapi untuk audit kinerja internal, penilaian akreditasi institusi, dan perencanaan strategis sumber daya manusia akademik. Hasil clustering langsung terhubung dengan indikator kinerja utama (KPI) dan standar mutu pendidikan tinggi yang diakui secara nasional.
8	Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan	1.Bernadus Gunawan Sudarsono 2.Sri Poedji Lestari. Media publikasi	Tujuan penelitian ini adalah untuk melakukan pengelompokan penerima beasiswa berdasarkan nilai yang	Kesimpulan dari penelitian ini adalah bahwa penggunaan metode clustering dengan algoritma K-Means efektif dalam mengelompokkan data penerima beasiswa. Proses	Saran yang diberikan dalam penelitian ini mencakup perlunya pengembangan lebih lanjut dalam penerapan algoritma K-Means untuk	Penelitian [12] 5. Fokus: **Alokasi sumber daya keuangan** (beasiswa) untuk mahasiswa. 6. Metode: K-Means clustering berdasarkan kriteria akademik dan prestasi.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Metode K-Means	adalah Jurnal Media Informatika Budidarma, dan tahun publikasinya adalah 2021	diakumulasikan. Dengan menggunakan metode clustering, penelitian ini bertujuan untuk memberikan beasiswa dengan jumlah dan besaran yang berbeda kepada penerima, mengingat bahwa beasiswa dari yayasan terbatas dan memiliki tingkatan terhadap pembagiannya	pengujian menunjukkan bahwa setelah iterasi pertama, terbentuk empat kelompok berdasarkan pembobotan data yang telah diakumulasikan. Pada iterasi kedua, data dihitung kembali untuk melihat jarak kedekatan antara satu data dengan data lainnya. Hasil akhir menunjukkan bahwa semua data memiliki kedekatan yang sama, sehingga terbentuk satu pengelompokan berdasarkan jarak kedekatan dengan nilai data	meningkatkan akurasi pengelompokan data penerima beasiswa. Peneliti juga menyarankan agar metode ini dapat diterapkan pada data yang lebih besar dan beragam untuk mendapatkan hasil yang lebih representatif. Kelemahan yang diidentifikasi dalam penelitian ini adalah bahwa algoritma K-Means sangat bergantung pada pemilihan jumlah cluster (k) yang tepat. Jika jumlah cluster yang ditentukan tidak sesuai, hasil pengelompokan dapat menjadi kurang akurat. Selain itu, K-Means juga sensitif terhadap outlier, yang dapat mempengaruhi	7. Tujuan: Distribusi beasiswa yang adil dan efisien sesuai kemampuan Yayasan. Sedangkan Penelitian pada Tesis ini: 5. Fokus: Evaluasi kinerja dan produktivitas dosen dalam tridharma perguruan tinggi. 6. Metode: K-Means clustering berdasarkan output penelitian dan pengabdian masyarakat. 7. Perbedaan Utama: Penelitian ini berfokus pada evaluasi kinerja sumber daya manusia akademik, bukan alokasi dana. Data yang digunakan adalah output kerja nyata (publikasi, kegiatan PKM) yang tercatat dalam sistem resmi, bukan data seleksi atau kriteria penerimaan. Hasil *clustering* ditujukan untuk perencanaan strategis institusi, penjaminan mutu, dan pemenuhan standar akreditasi nasional, sehingga memiliki dampak pada kebijakan dan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					pusat cluster dan hasil akhir pengelompokan	reputasi perguruan tinggi secara menyeluruh.

2.3. Landasan Teori

Penggunaan sistem informasi berbasis web dalam pengelolaan Beban Kerja Dosen (BKD) memiliki potensi besar untuk mempercepat proses administrasi, meningkatkan akurasi data, dan memungkinkan pemantauan yang lebih efektif terhadap kinerja dosen [4].

2.3.1. Data Mining

Data mining merupakan sebuah proses ekstraksi yang bertujuan untuk menggali informasi yang mungkin belum diketahui secara eksplisit, namun tersembunyi dalam kumpulan data. Proses ini umumnya digunakan untuk mengidentifikasi pola, tren, dan hubungan dalam data yang besar dan kompleks. Data mining memiliki keterkaitan erat dengan berbagai bidang, termasuk statistik, algoritma pola, teknologi basis data, pembelajaran mesin, algoritma komputasi, dan komputasi berkinerja tinggi. Melalui penggabungan konsep-konsep dari berbagai bidang ini, data mining memungkinkan pengguna untuk mengambil wawasan yang berharga dari data yang ada, sehingga dapat digunakan untuk pengambilan keputusan yang lebih baik dalam berbagai konteks [18].

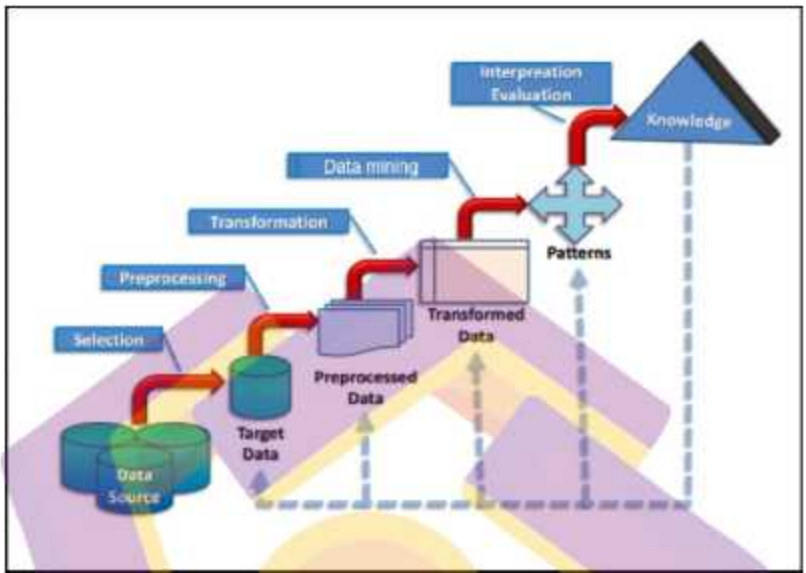
Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Database* (KDD) [19]

Gambar 2. 1 Ketertarikan data mining

Selain itu, dalam data mining terdapat berbagai metode yang digunakan untuk menganalisis dan menggali informasi dari kumpulan data.

2.3.1.1 Knowledge Discovery Database (KDD)

Data mining dan *Knowledge Discovery in Databases* (KDD) sering digunakan secara bergantian untuk menggambarkan proses menemukan informasi tersembunyi dalam basis data yang besar. *Knowledge Discovery in Databases* (KDD) adalah proses yang mengubah data mentah menjadi informasi yang berguna. Proses KDD ini terdiri dari serangkaian langkah yang dilakukan secara berurutan [20].



Gambar 2.2 Langkah Knowledge Discovery in Databases (KDD) (A. Villanueva, 2020)

Pangestu dan Noris menggambarkan langkah-langkah proses *Knowledge Discovery Database* (KDD) sebagai berikut [21]:

1) Data Selection

   Pemilihan data dari sekumpulan data operasional harus dilakukan sebelum memulai tahap penggalian informasi dalam KDD. Data yang telah dipilih kemudian akan digunakan untuk proses data mining.

2) Pre-Processing dan Data Cleaning

   Sebelum proses data mining dapat dilakukan, diperlukan langkah pembersihan data (data cleaning) pada data yang menjadi fokus KDD. Proses ini mencakup penghapusan duplikasi, pemeriksaan konsistensi data, serta perbaikan kesalahan yang ditemukan dalam data.

3) Transformation

Transformasi adalah proses mengubah data ke dalam bentuk yang sesuai untuk diolah dalam tahap data mining.

4) Data Mining

Data mining adalah proses untuk mencari pola atau informasi menarik dari data yang telah dipilih, dengan menggunakan teknik atau metode tertentu.

5) Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses data mining harus disajikan dalam bentuk yang mudah dipahami oleh pihak terkait, seperti melalui visualisasi atau tampilan yang dapat menjelaskan hasil sistem.

2.3.1.2 Tugas Utama Data Mining

Data Mining memiliki enam tugas utama yang dapat dilakukan, sebagai berikut [22] :

1) Deskripsi

Deskripsi adalah proses yang digunakan oleh peneliti dan analis untuk memahami dan menggambarkan pola serta kecenderungan yang ada dalam suatu kumpulan data. Tujuannya adalah untuk memberikan gambaran umum mengenai data tersebut, sehingga dapat dilihat lebih jelas bagaimana data tersebut terdistribusi atau bagaimana tren tertentu muncul di dalamnya. Deskripsi ini bisa melibatkan statistik sederhana, grafik, atau teknik visualisasi lainnya untuk menyajikan informasi secara intuitif.

2) Estimasi

Estimasi adalah proses yang mirip dengan klasifikasi, namun perbedaannya

terletak pada tipe variabel target yang digunakan. Dalam estimasi, variabel target bersifat numerik, sehingga tujuan dari estimasi adalah untuk memperkirakan nilai numerik dari data yang belum diketahui berdasarkan data yang ada. Misalnya, estimasi dapat digunakan untuk menentukan perkiraan pendapatan seseorang berdasarkan data demografis dan riwayat pekerjaan mereka.

3) Prediksi

Prediksi juga memiliki kesamaan dengan klasifikasi dan estimasi, namun yang membedakannya adalah fokus pada hasil yang akan terjadi di masa mendatang. Prediksi melibatkan penggunaan data historis untuk memperkirakan nilai atau kejadian yang akan datang. Contohnya, prediksi dapat digunakan untuk memproyeksikan penjualan produk pada bulan depan berdasarkan pola penjualan yang ada saat ini dan di masa lalu.

4) Klasifikasi

Klasifikasi adalah proses di mana data diorganisir ke dalam kategori atau kelas tertentu berdasarkan variabel target yang bersifat kategorikal. Tujuan dari klasifikasi adalah untuk menentukan kelompok mana data baru akan dimasukkan, berdasarkan pelajaran dari data sebelumnya. Misalnya, klasifikasi bisa digunakan untuk mengategorikan email sebagai spam atau bukan spam.

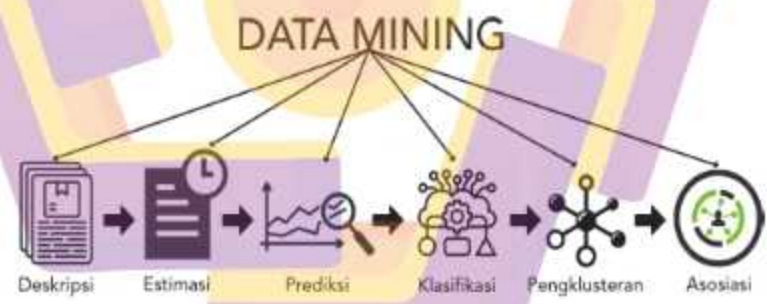
5) Clustering

Clustering adalah teknik yang digunakan untuk mengidentifikasi dan mengelompokkan data yang memiliki kesamaan atau kemiripan karakteristik. Metode ini berbeda dari klasifikasi karena tidak menggunakan variabel target

dan bersifat tanpa arahan (unsupervised). Dalam Clustering, algoritma secara otomatis mengelompokkan data ke dalam cluster berdasarkan tingkat kesamaan antar data, sehingga pola yang tersembunyi dalam data dapat ditemukan tanpa pengawasan eksplisit.

6) Asosiasi

Asosiasi adalah proses dalam data mining yang bertujuan untuk menemukan hubungan antara atribut-atribut yang muncul bersama-sama dalam suatu waktu atau kejadian tertentu. Contoh umum dari asosiasi adalah analisis keranjang belanja, di mana ditemukan pola bahwa pelanggan yang membeli satu produk tertentu sering kali juga membeli produk lain pada saat yang sama. Teknik ini berguna untuk memahami keterkaitan antar item dalam data yang besar dan kompleks.



Gambar 2.3 Tugas Data Mining

2.3.2. Clustering

2.3.2.1 Pengertian *Clustering*

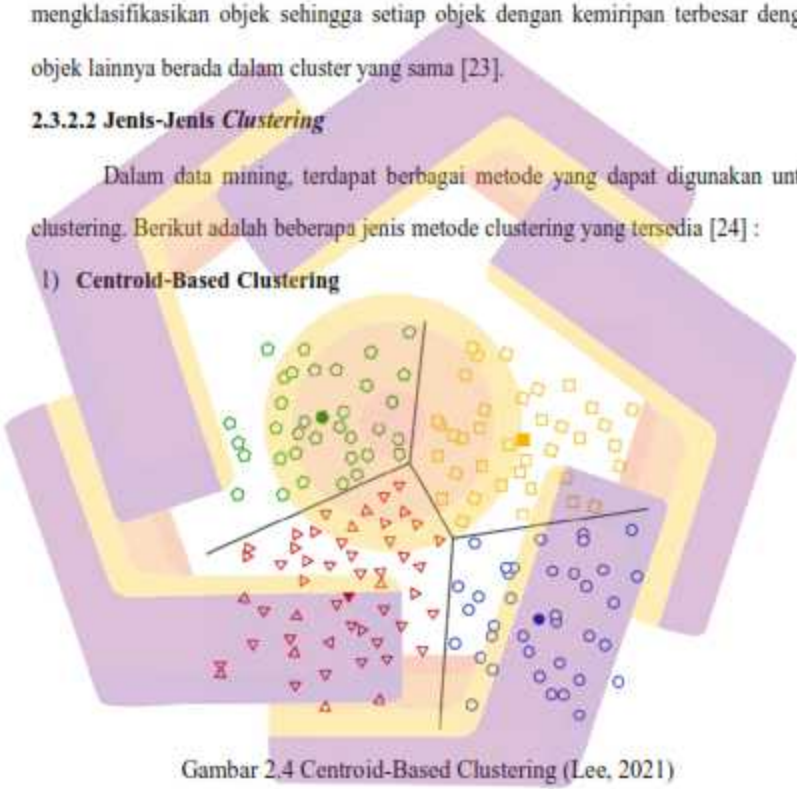
Clustering adalah salah satu bagian dari data mining yang melibatkan proses pengelompokan sampel yang serupa ke dalam kelompok-kelompok yang dikenal

sebagai cluster. Dalam setiap cluster, anggota-anggota memiliki kesamaan yang tinggi di antara mereka dan perbedaan yang jelas dibandingkan dengan anggota dari cluster lain. Analisis cluster adalah teknik multivariat yang bertujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki. Teknik ini mengklasifikasikan objek sehingga setiap objek dengan kemiripan terbesar dengan objek lainnya berada dalam cluster yang sama [23].

2.3.2.2 Jenis-Jenis *Clustering*

Dalam data mining, terdapat berbagai metode yang dapat digunakan untuk clustering. Berikut adalah beberapa jenis metode clustering yang tersedia [24] :

1) **Centroid-Based Clustering**

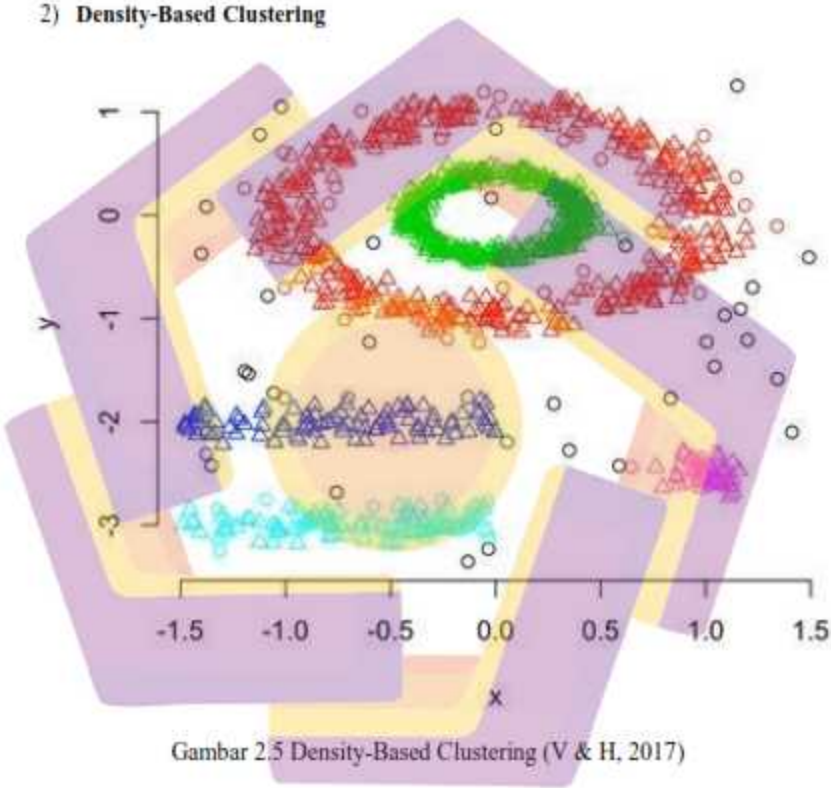


Gambar 2.4 Centroid-Based Clustering (Lee, 2021)

Centroid-Based Clustering adalah metode clustering yang mengelompokkan data ke dalam cluster non-hierarkis. Metode ini sangat sensitif terhadap outlier dan menggunakan algoritma iteratif dalam proses clustering-nya. Dalam pendekatan ini, sebuah *cluster* dibentuk berdasarkan jarak terdekat

antara titik data dengan pusat *cluster* (centroid). Pusat centroid ditentukan dengan cara meminimalkan jarak antara titik data dan pusat cluster tersebut. Algoritma yang paling terkenal dalam metode *Centroid-Based Clustering* adalah K-Means Clustering.

2) **Density-Based Clustering**

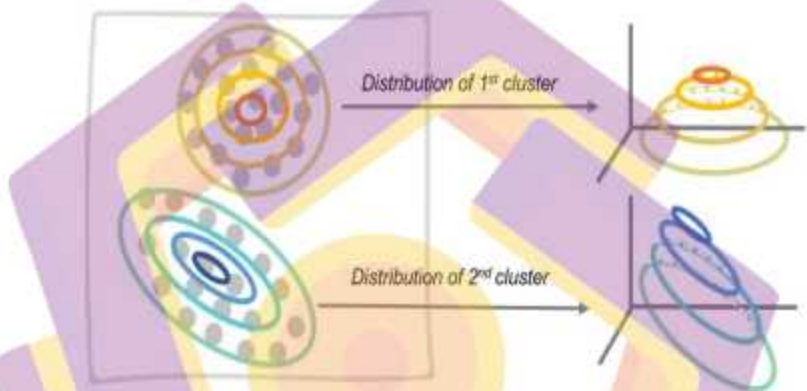


Gambar 2.5 Density-Based Clustering (V & H, 2017)

Density-Based Clustering adalah metode yang mengelompokkan area dengan kepadatan data yang sama ke dalam satu kelompok. Dalam metode ini, cluster dibentuk berdasarkan kepadatan dari setiap titik data. Area yang padat atau memiliki banyak data akan dianggap sebagai satu cluster, sementara area dengan kepadatan rendah akan dianggap sebagai outlier. Algoritma yang menggunakan

pendekatan Density-Based Clustering meliputi DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points to Identify Clustering Structure), dan HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise).

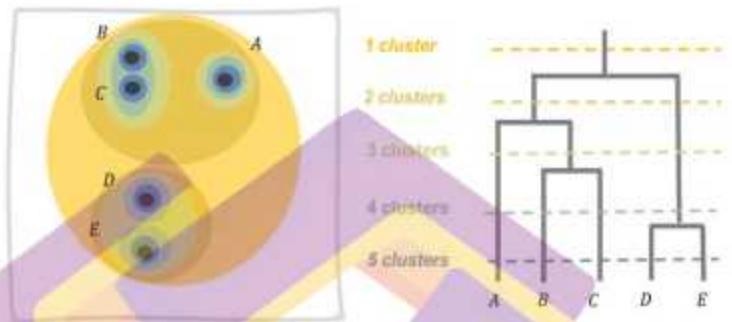
3) Distribution-Based Clustering



Gambar 2.6 Distribution-Based Clustering (Gao et al., 2023)

Distribution-Based Clustering adalah metode yang mengasumsikan bahwa data terdiri dari distribusi tertentu. Seiring dengan meningkatnya jarak dari pusat distribusi, probabilitas bahwa suatu titik termasuk ke dalam kelompok distribusi tersebut akan menurun. Metode ini sangat efektif untuk data sintetis dan cluster dengan ukuran yang bervariasi. Algoritma yang umum digunakan dalam Distribution-Based Clustering adalah Expectation-Maximization.

4) Hierarchical Clustering



Gambar 2.7 Hierarchical Clustering (Gao et al., 2023)

Hierarchical Clustering adalah metode yang serupa dengan Centroid-Based Clustering, karena mengelompokkan data berdasarkan jarak terdekat antara titik data. Pada dasarnya, titik data yang lebih dekat cenderung memiliki karakteristik yang lebih mirip dibandingkan dengan titik data yang lebih jauh. Hasil pengelompokan ini biasanya direpresentasikan menggunakan dendogram.

2.3.3. K-Means

K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric. Algoritma K-Means termasuk *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma K-Means, setiap data harus termasuk ke cluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk cluster

tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster lainnya [25].

Algoritma *K-Means* merupakan teknik analisis data yang menggunakan sistem partisi untuk melakukan proses pengelompokan data. Suatu data dikelompokkan ke dalam satu *cluster* berdasarkan kemiripan atribut yang dimiliki. Kemiripan ini bisa diketahui dengan mengukur jarak setiap data dengan pusat *cluster* (*centroid*). Metode *K-Means Clustering* merupakan salah satu metode *Clustering* untuk mengelompokkan data yang memiliki jumlah data besar dan proses yang cepat dan efisien [9]. Berikut adalah langkah perhitungan algoritma *K-Means*

1) Tentukan berapa banyak jumlah *cluster* (K) atau kelompok
2) Tentukan secara acak pusat *cluster* awal (*centroid*)
3) Ukur jarak setiap data dengan pusat *cluster* (*centroid*) yaitu dengan rumus Euclidean Distance pada rumus berikut.

$$D_{i,j} = \sqrt{(x_{1p} - x_{pq})^2 + (x_{2p} - x_{2q})^2 + \ldots + (x_{rp} - x_{rq})^2}$$

Keterangan :

$D(\_p,\_q)\_\_$ = jarak data ke-p dengan pusat *cluster* q

$X(\_r,\_p)\_\_$ = data ke-p pada atribut data ke-r

$X(\_r,\_q)\_\_$ = titik pusat ke-q pada atribut

4) Kelompokkan setiap data ke dalam *cluster* berdasarkan jarak minimum
5) Lakukan proses iterasi dengan menentukan pusat cluster (*centroid*) baru dengan rumus pada rumus berikut.

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

Keterangan :

v = *centroid* pada *cluster*

x(_*p*)_= objek ke-p

n= banyaknya objek/jumlah

6) Ulangi langkah 3 hingga tidak terdapat perubahan *cluster* pada setiap data dari proses iterasi sebelumnya.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini bertujuan untuk menganalisis laporan Beban Kerja Dosen (BKD) pada bidang penelitian dan Pengabdian kepada Masyarakat (PkM) di Universitas Islam Sultan Aji Muhammad Idris Samarinda dengan menggunakan metode *clustering*, khususnya K-Means clustering, untuk membantu memenuhi kriteria akreditasi BAN-PT APT dan mencapai *Key Performance Indicators* (KPI) atau Indikator Kinerja Utama (IKU). Pada bab ini, dijelaskan secara terperinci jenis, sifat dan desain pendekatan penelitian, metode pengumpulan data, metode analisis data, serta alur penelitian yang digunakan dalam penelitian ini.

Jenis penelitian ini dikategorikan sebagai penelitian terapan (*applied research)* karena berfokus pada pemecahan masalah praktis dalam konteks nyata, yaitu analisis produktivitas dosen untuk mendukung pengambilan keputusan strategis dalam evaluasi kinerja dan dalam penjaminan mutu perguruan tinggi.

Sifat penelitian bersifat eksplanatori-analitik yang menggabungkan pendekatan eksploratif untuk menemukan pola tersembunyi dalam data dengan analisis sistematis untuk menjelaskan hubungan antar variabel produktivitas. Penelitian tidak hanya mendeskripsikan data secara numerik, tetapi juga menjelaskan pengelompokan pola produktivitas dosen melalui teknik *clustering*.

Pendekatan penelitian menggunakan paradigma kuantitatif dengan metodologi *Knowledge Discovery in Databases* (KDD) yang terstruktur dalam

tahapan: (1) seleksi data dari sistem e-BKD, (2) *preprocessing* dan *cleaning* data, (3) transformasi dan preparasi data, (4) *data mining* dengan implementasi algoritma K-Means *clustering*, serta (5) interpretasi dan evaluasi hasil untuk pengambilan keputusan. Pendekatan ini memungkinkan ekstraksi pengetahuan dari data historis BKD untuk identifikasi pola produktivitas yang relevan dengan target akreditasi dan KPI institusi.

3.2. Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan secara sistematis, di mana data dikategorikan menjadi data primer yang bersumber dari sistem e-BKD dan data sekunder yang berasal dari dokumen kebijakan institusi serta literatur akademik.

3.2.1 Data untuk menganalisis Laporan BKD

1) Data yang Dibutuhkan yakni:
   a. Data historis Beban Kerja Dosen (BKD) bidang Penelitian (id_bidang = 2) dan Pengabdian kepada Masyarakat (PKM) (id_bidang = 3)
   b. Atribut utama: identitas dosen, jenis kegiatan, satuan kredit semester (SKS), periode akademik, bukti kinerja, status validasi
   c. Periode: 6 semester (ganjil 2021/2022 sampai genap 2023/2024)
2) Sumber Data yang dibutuhkan yakni:
   a. Database aplikasi e-BKD Lembaga Penjaminan Mutu (LPM) UIN Sultan Aji Muhammad Idris Samarinda
   b. Tabel app_kegiatan sebagai sumber data utama (struktur seperti Tabel 3.1)

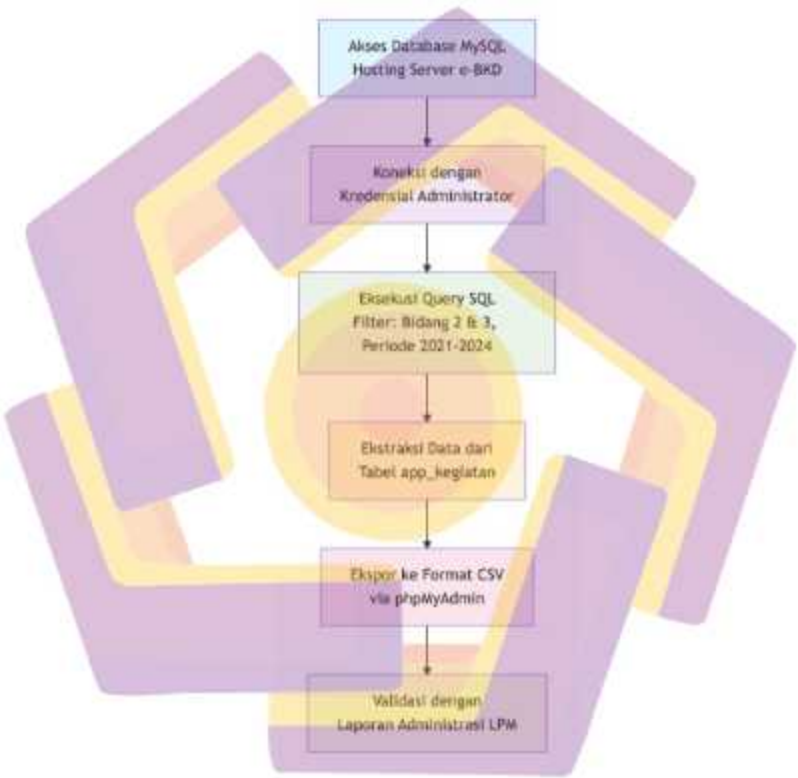
Tabel 3.1 Struktur Data pada Tabel app_kegiatan

id_user	id_bidang	jenis_kegiatan	id_pelaksanaan	id_tahun_akademik
73	2	Penelitian	18	2
73	1	Mengajar Metode dan Pendekatan Studi Islam	1	2
73	1	Mengajar Matkul Tafsir Tematis 2 (Lingkungan Hidup) IAT/41	1	2
80	1	Mengajar MK Bahasa Indonesia Keilmua, Prodi KPI, 1 Lokal, Semester 1	1	1
68	1	Mengajar MK Perkembangan Pemikiran Modern Dalam Islam Lokal MD31	1	1
98	1	Perkuliahan Pada Mata Kuliah Pancasila PAI 2	1	1
113	1	Mengajar Mata Kuliah Pengembangan Bahan Ajar 1 / PAI 6	1	1
202	1	Mengajar Telaah dan Pengembangan Kurikulum	1	1
69	1	Mengajar Mata Kuliah Pemahaman Individu Tes dan Non Tes, Prodi BKI 51, Smt. V, SKS 3	1	1

3) Cara Pengambilan:

Pengambilan data dilakukan melalui akses langsung ke database MySQL pada *server hosting* sistem e-BKD. Menggunakan kredensial administrator, peneliti melakukan koneksi ke database dan mengeksekusi *query* SQL terstruktur untuk mengekstrak data dari tabel app_kegiatan. *Query* difilter untuk bidang Penelitian (id_bidang=2) dan PkM (id_bidang=3) dengan periode semester ganjil 2021/2022 hingga semester genap 2023/2024. Hasil *query* diekspor ke format CSV melalui antarmuka phpMyAdmin untuk proses analisis lebih lanjut menggunakan Python dan *library data science*. Validasi kelengkapan data dilakukan dengan

membandingkan jumlah *records* yang diekstrak dengan laporan administrasi LPM UIN Sultan Aji Muhammad Idris Samarinda. Proses pengambilan data divisualisasikan dalam diagram alir berikut:



Gambar 3.1 Diagram Alir Pengambilan Data dari Sistem e-BKD

3.3. Metode Analisis Data

Data yang telah dikumpulkan dianalisis dengan menggunakan metode *K-Means clustering*. Tahapan analisis data meliputi:

1) Preprocessing Data

Tahap *preprocessing* bertujuan untuk menyiapkan data agar siap digunakan dalam proses clustering. Tahapan ini mengacu pada alur kerja yang ditunjukkan pada Gambar 3.2 dan mencakup beberapa sub-tahapan sebagai berikut:

a. Pengambilan Data

Data dikumpulkan dari sumber yang telah ditentukan, mencakup atribut-atribut utama seperti nama dosen, bidang penelitian, bidang PkM, jenis publikasi, dan indikator kinerja.

b. Pembersihan Data (Data Cleaning)

1) Pengisian nilai hilang (*missing value handling*) menggunakan metode imputasi atau penghapusan jika diperlukan.
2) Penghapusan data duplikat untuk menghindari bias dalam analisis.
3) Koreksi kesalahan pada entri data, seperti kesalahan penulisan atau inkonsistensi format.

c. Transformasi Data

Transformasi dilakukan agar data dapat diproses secara matematis dalam algoritma K-Means. Langkah-langkahnya meliputi:

1) Konversi data kategorikal ke dalam bentuk numerik (misalnya: bidang penelitian, jenis publikasi) menggunakan teknik *label encoding* atau *one-hot encoding*.

2) Normalisasi data untuk menyamakan skala antar atribut numerik, terutama pada indikator kinerja yang memiliki rentang nilai berbeda.
3) Pengubahan format tanggal ke dalam format yang seragam, serta bila diperlukan dikonversi menjadi nilai numerik (misalnya semester dalam bentuk angka).

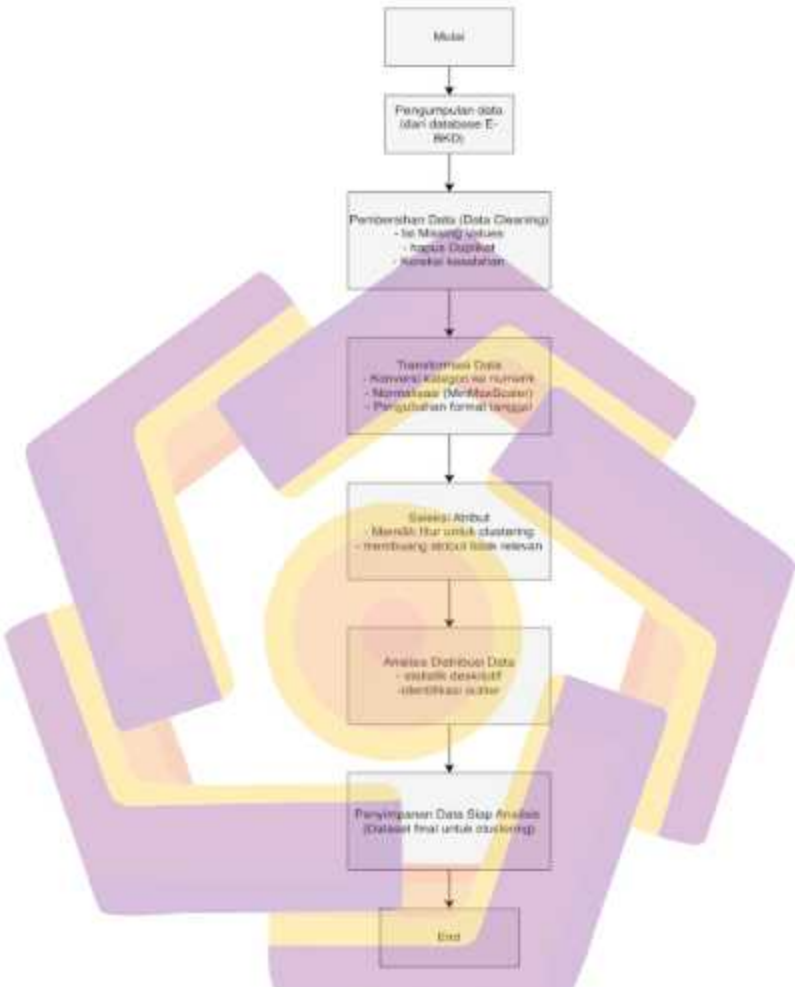
d. Seleksi Atribut

Atribut yang relevan untuk proses *clustering* dipilih berdasarkan tujuan analisis. Atribut yang digunakan meliputi bidang penelitian, bidang PkM, jenis publikasi, serta indikator kinerja. Atribut seperti nama dosen tidak diikutsertakan dalam proses perhitungan jarak, namun digunakan sebagai pelengkap dalam interpretasi hasil klaster.

e. Pengelompokan Awal dan Penyimpanan Data Siap Analisis

Data yang telah melalui proses pembersihan dan transformasi kemudian dikategorikan sementara berdasarkan semester atau indikator kinerja tertentu untuk memudahkan analisis lanjutan. Seluruh data yang telah siap digunakan selanjutnya disimpan dalam format yang terstruktur sebagai masukan untuk proses clustering.

Sebelum ke tahap pemodelan, dilakukan analisis distribusi data untuk memahami karakteristik awal data, termasuk sebaran nilai, potensi pencilan (*outlier*), serta pola awal yang mungkin muncul. Gambar 3.2 berikut menunjukkan alur lengkap dari tahapan preprocessing yang dilaksanakan.



Gambar 3.2 Alur Preprocessing data

2) Penanganan Outlier dan Justifikasi Penggunaan K-Means

Deteksi Outlier: Menggunakan metode *Interquartile Range* (IQR) dengan ambang batas 1,5×IQR pada kelima fitur clustering. Hasil deteksi menunjukkan bahwa dari 214 dosen, terdapat 37 outlier (17,3%) yang

sebagian besar merupakan dosen dengan produktivitas sangat tinggi (total SKS > 50). Outlier tidak dihapus dari dataset karena mencerminkan variasi nyata kinerja dosen super produktif dan menghapusnya justru akan mengurangi representasi populasi. Namun, dampak outlier terhadap metrik *Davies-Bouldin Index* (DBI) dicatat sebagai cerminan *messiness data* (kualitas data yang belum ideal).

Justifikasi Kesesuaian Data dengan K-Means : K-Means mengasumsikan cluster berbentuk *spherical* (globular) dan berukuran relatif sama. Untuk menguji kesesuaian, dilakukan analisis *Principal Component Analysis* (PCA) pada data yang telah dinormalisasi. Hasilnya menunjukkan bahwa dua komponen utama mampu menjelaskan 82,1% total varians data (PC1 = 49,5%, PC2 = 32,6%), yang mengindikasikan struktur data tidak terlalu kompleks dan cenderung *globular*. Dengan demikian, data BKD yang telah diproses memenuhi asumsi dasar K-Means.

3) Proses K-Means Clustering

Setelah data siap, proses clustering dilakukan menggunakan algoritma K-Means. Proses ini meliputi:

a. Penentuan jumlah cluster.
b. Menentukan centroid awal.
c. Menghitung jarak data ke masing-masing centroid.
d. Mengelompokkan data berdasarkan jarak terdekat ke centroid.
e. Mengulangi proses hingga tidak ada perubahan posisi centroid.

Gambar 3.3 Flowchart K-Means

Hasil clustering digunakan untuk mengukur ketercapaian Key Performance Indicators (KPI) atau Indikator Kinerja Utama UINSI Samarinda dalam bidang Penelitian dan Pengabdian kepada Masyarakat (PkM) dan juga untuk pemenuhan kriteria akreditasi BAN-PT Akreditasi Perguruan Tinggi.

1) Key Performance Indicators (KPI) atau Indikatator Kinerja Utama UIN Sultan Aji Muhammad Idris Samarinda.
   a. Publikasi Ilmiah Penelitian dan PkM

Tabel 3.2 Publikasi Ilmiah Penelitian dan PkM UIN Sultan Aji Muhammad Idris Samarinda

No	KPI atau IKU	Target Pencapaian
1	Jumlah publikasi di jurnal nasional tidak terakreditasi	14% jumlah dosen tetap
2	Jumlah publikasi di jurnal nasional terakreditasi	13% jumlah dosen tetap
3	Jumlah publikasi di jurnal internasional	12% jumlah dosen tetap
4	Jumlah publikasi di jurnal internasional bereputasi	10% jumlah dosen tetap
5	Jumlah publikasi di jurnal internasional bereputasi tinggi	≥ 12% jumlah dosen tetap

b. Produktivitas Penelitian dan PKM Dosen

Tabel 3.3 Persentase dan Rata-rata Dana Penelitian dan PkM Dosen UIN Sultan Aji Muhammad Idris Samarinda

No	Indikator	Persentase/Target Pencapaian
1	Persentase jumlah penelitian terhadap jumlah keseluruhan dosen berdasarkan sumber biaya penelitian	1) Luar Negeri: ≥ 0,1%
		2) Dalam Negeri (Luar PT): ≥ 0,16%
		3) PT/Mandiri: ≥ 0,2%
2	Persentase jumlah penelitian terhadap jumlah keseluruhan dosen berdasarkan sumber biaya penelitian	1) Luar Negeri: ≥ 0,2%
		2) Dalam Negeri (Luar PT): ≥ 0,18%
		3) PT/Mandiri: ≥ 0,23%
3	Persentase jumlah PkM terhadap jumlah keseluruhan dosen berdasarkan sumber biaya penelitian	1) Luar Negeri: ≥ 0,05%
		2) Dalam Negeri (Luar PT): ≥ 0,08%
		3) PT/Mandiri: ≥ 0,14%
4	Persentase jumlah PkM terhadap jumlah keseluruhan dosen berdasarkan sumber biaya penelitian	1) Luar Negeri: ≥ 0,07%
		2) Dalam Negeri (Luar PT): ≥ 0,1%
		3) PT/Mandiri: ≥ 0,16%
5	Rata-rata dana penelitian dosen per tahun	≥ 20.000.000 Rupiah
6	Rata-rata dana PkM dosen per tahun	≥ 5.000.000 Rupiah

2) Matrik Akreditasi BAN-PT Akreditasi perguruan Tinggi

a. Publikasi Ilmiah Penelitian dan PkM

Tabel 3.4 Matrik BAN-PT. Penelitian dan Jumlah Publikasi

C.9.4.B) Penelitian dan jumlah publikasi di jurnal dalam 3 tahun terakhir		
Kode	Keterangan	Jumlah
NA_1	Jumlah publikasi di jurnal nasional tidak terakreditasi	0
NA_2	Jumlah publikasi di jurnal nasional terakreditasi	0
NA_3	Jumlah publikasi di jurnal internasional	0
NA_4	Jumlah publikasi di jurnal internasional bereputasi	66
NDT	Jumlah dosen tetap	214
R_a	$R_a = NA_4 / NDT$	0,27
R_n	$R_n = (NA_2 + NA_3) / NDT$	0,00
R_t	$R_t = NA_1 / NDT$	0,00
	Skor	4,00

b. Produktivitas Penelitian dan PkM Dosen

Berikut ini merupakan rata-rata produktivitas penelitian Dosen UIN Sultan Aji Muhammad Idris samarinda.

Tabel 3.5 Rata-rata Penelitian

Rata-rata penelitian/dosen/tahun dalam 3 tahun terakhir		
Kode	Keterangan	Jumlah
N_1	Jumlah penelitian dengan biaya luar negeri	100
N_n	Jumlah penelitian dengan biaya dalam negeri di luar PT	100
N_l	Jumlah penelitian dengan biaya dari PT atau mandiri	500
NDT	Jumlah dosen tetap	214
R_l	$R_l = N_1 / 3 / NDT$	0,14
R_n	$R_n = N_n / 3 / NDT$	0,14
R_l	$R_l = N_l / 3 / NDT$	0,68
	Skor	4,00

Tabel 3.6 Rata-rata PKM

Rata-rata PkM/dosen/tahun dalam 3 tahun terakhir		
Kode	Keterangan	Jumlah
N_t	Jumlah PkM dengan biaya luar negeri	0
N_n	Jumlah PkM dengan biaya dalam negeri di luar PT	0
N_l	Jumlah PkM dengan biaya dari PT atau mandiri	100
NDT	Jumlah dosen tetap	245
R_t	$R_t = N_t / 3 / NDT$	0,00
R_n	$R_n = N_n / 3 / NDT$	0,00
R_l	$R_l = N_l / 3 / NDT$	0,14
	Skor	0,27

Tabel 3. 7 Dana Penelitian

Rata-rata dana penelitian dosen/tahun		
Kode	Keterangan	Jumlah
D_p	Jumlah dana penelitian yang diperoleh dosen tetap dalam 3 tahun terakhir	550.000.000
NDT	Jumlah dosen tetap	214
D_pDT	$D_p / 3 / NDT$	748.299
	Skor	0,15

Tabel 3.8 Dana PKM

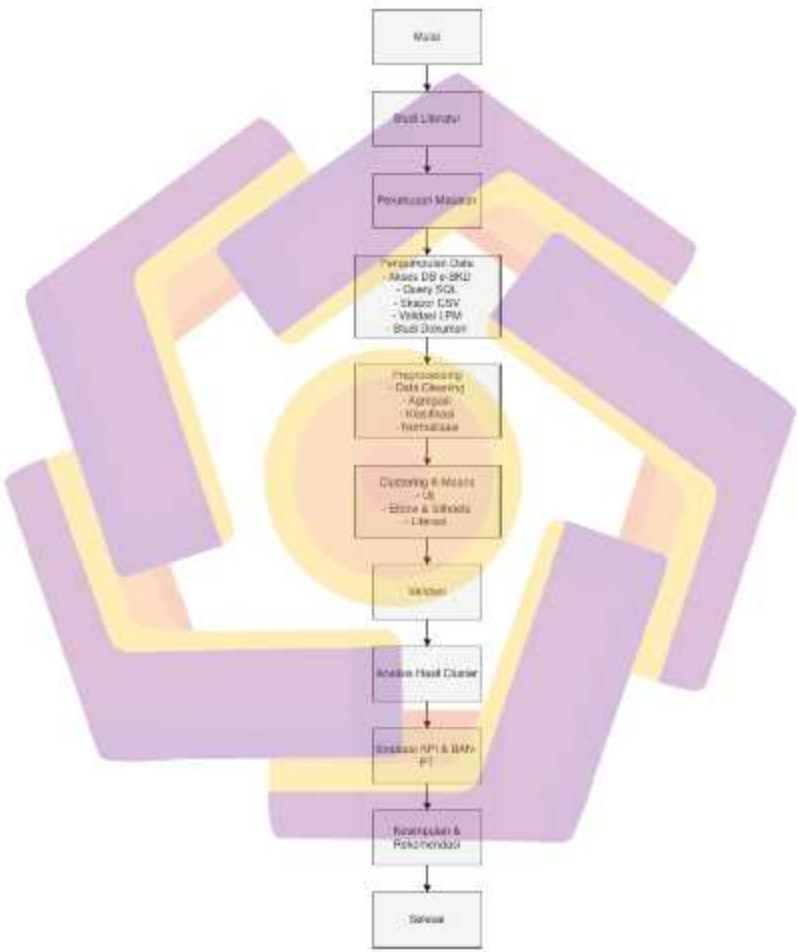
Rata-rata dana PkM dosen/tahun		
Kode	Keterangan	Jumlah
D_{pkm}	Jumlah dana PkM yang diperoleh dosen tetap dalam 3 tahun terakhir	720.000.000
NDT	Jumlah dosen tetap	214
$D_{pkm}DT$	$D_{pkm} / 3 / NDT$	979.592
	Skor	0,78

c. Analisis Hasil Clustering

Setelah pengelompokan data selesai, hasil clustering dianalisis untuk menentukan pola atau tren yang muncul dalam beban kerja dosen. Analisis ini akan mencakup interpretasi dari hasil clustering untuk setiap kelompok, serta evaluasi capaian KPI atau IKU dan pemenuhan kriteria akreditasi BAN-PT Akreditasi Perguruan Tinggi (APT) berdasarkan hasil pengelompokan yang dilakukan.

3.4. Alur Penelitian

Gambar 3.4 menyajikan diagram alir penelitian yang menggambarkan seluruh tahapan secara sistematis dari awal hingga akhir.



Gambar 3.4 Diagram Alir Penelitian

Berdasarkan diagram alir pada Gambar 3.4, penelitian ini dilakukan dalam tiga tahapan utama. Tahap pertama adalah studi literatur untuk membangun

landasan teori dan mengidentifikasi *research gap*, kemudian dilanjutkan dengan perumusan masalah penelitian. Tahap kedua adalah pengumpulan data primer melalui akses langsung ke database MySQL sistem e-BKD pada *server hosting*. Data diekstraksi menggunakan *query* SQL dengan filter bidang Penelitian (id_bidang=2) dan PKM (id_bidang=3) untuk periode semester ganjil 2021/2022 hingga semester genap 2023/2024. Hasil *query* diekspor ke format CSV dari tabel app_kegiatan dan selanjutnya divalidasi bersama pihak LPM UINSI Samarinda. Selain itu, data sekunder berupa dokumen KPI/IKU dan matriks akreditasi BAN-PT APT juga dikumpulkan melalui studi dokumen.

Tahap ketiga adalah analisis data dengan pendekatan *Knowledge Discovery in Databases* (KDD). Proses diawali dengan *preprocessing* data yang mencakup pembersihan data dengan menghapus kolom yang 100% kosong dan mengisi *missing values* pada kolom SKS dengan nilai 0, normalisasi nama dosen, agregasi data per dosen hingga menghasilkan 214 profil dosen, klasifikasi jenis publikasi menggunakan metode *keyword matching*, serta normalisasi fitur menggunakan MinMaxScaler. Deteksi outlier menggunakan metode IQR mengidentifikasi 37 outlier (17,3%) yang tidak dihapus karena mencerminkan variasi nyata kinerja dosen super produktif. Setelah data siap, dilakukan proses *clustering* menggunakan algoritma K-Means dengan pengujian jumlah *cluster* dari K=2 hingga K=7.

Penentuan jumlah *cluster* optimal menggunakan *Elbow Method* dan *Silhouette Score* menghasilkan K=2 dengan *Silhouette Score* 0,3882 yang termasuk dalam kategori cukup baik (0,3–0,5). Validasi hasil *clustering* kemudian dilakukan

dengan *Silhouette Score*, *Davies-Bouldin Index*, dan visualisasi PCA. Tahap keempat adalah interpretasi hasil dan penarikan kesimpulan. Analisis profil *cluster* berhasil mengidentifikasi dua kelompok produktivitas dosen, yaitu Cluster 2 (Produktivitas Tinggi) sebanyak 54 dosen (25,2%) dengan rata-rata total SKS 27,07, dan Cluster 1 (Produktivitas Sedang/Rendah) sebanyak 160 dosen (74,8%) dengan rata-rata total SKS 12,37. Uji beda Mann-Whitney menunjukkan perbedaan signifikan antar *cluster* (p-value = 0,000). Gap produktivitas total SKS antara kedua *cluster* mencapai 2,19 kali lipat, dengan *cluster* tinggi menyumbang hampir seluruh publikasi internasional bereputasi. Evaluasi pencapaian KPI menunjukkan bahwa empat dari lima indikator tercapai, namun indikator publikasi internasional (80,8%) dan internasional bereputasi tinggi (88,5%) belum memenuhi target. Proyeksi skor akreditasi BAN-PT APT adalah 84,6 dari 400 dengan predikat C (Baik). Tahap akhir penelitian ini adalah penarikan kesimpulan dan penyusunan rekomendasi. Kesimpulan utama membuktikan bahwa metode *clustering* efektif untuk menganalisis laporan beban kerja dosen pada bidang penelitian dan PkM, meskipun terbatas oleh *data quality issue* (81,5% aktivitas tidak terklasifikasi). Berdasarkan hasil tersebut, dirumuskan rekomendasi strategis berbasis *cluster* yang dituangkan dalam bentuk *roadmap* tiga tahun (2024-2027).

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1 Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini bersumber dari aplikasi e-BKD Lembaga Penjaminan Mutu (LPM) UIN Sultan Aji Muhammad Idris Samarinda. Perlu ditekankan bahwa data yang diekstraksi hanya mencakup aktivitas pada bidang Penelitian (id_bidang = 2) dan Pengabdian kepada Masyarakat (id_bidang = 3). Komponen BKD lainnya, yaitu Pendidikan dan Pengajaran serta Penunjang, tidak disertakan karena berada di luar fokus penelitian. Dengan demikian, seluruh analisis dan interpretasi hasil hanya berlaku untuk kontribusi dosen pada bidang penelitian dan PkM, bukan total beban kerja dosen secara keseluruhan.

Tabel 4.1 Ruang lingkup data BKD yang Dianalisis

Komponen BKD	Id Bidang	Cakupan Dalam Penelitian	Keterangan
Pendidikan & Pengajaran	1	Tidak Dianalisis	Di luar fokus penelitian
Penelitian	2	Dianalisis	Fokus utama penelitian
Pengabdian (PkM)	3	Dianalisis	Fokus utama penelitian
Penunjang	4	Tidak Dianalisis	Di luar fokus penelitian

Berikut ini merupakan Gambaran Umum Dataset di UIN Sultan Aji Muhammad Samarinda:

1) Total records aktivitas: 2.762 aktivitas penelitian dan PKM
2) Jumlah data user dosen: 214 dosen
3) Periode data: 3 tahun akademik (2021/2022 sampai 2023/2024)

4) Bidang aktivitas: Penelitian (id_bidang = 2) sebanyak 1.098 aktivitas dan PKM (id_bidang = 3) sebanyak 1.664 aktivitas
5) Memory usage: 3.53 MB

Berikuti ini merupakan Distribusi Temporal aktivitas Dosen di UIN Sultan Aji Muhammad Samarinda:

Tabel 4.2 Distribusi Temporal Aktivitas Dosen

Tahun Akademik	Jumlah Aktivitas	Persentase
2021/2022	915	33.1%
2022/2023	1191	43.1%
2023/2024	656	23.8%

Dataset mentah yang diperoleh dari database e-BKD dalam format CSV kemudian diimpor menggunakan Python dengan library pandas. Proses impor dilakukan melalui fungsi pd.read_csv() sehingga seluruh record termuat ke dalam DataFrame untuk memudahkan eksplorasi dan manipulasi data.

Terjadi peningkatan aktivitas pada tahun 2022/2023 (+30.2%), namun menurun pada 2023/2024 (-44.9%). Penurunan ini kemungkinan karena data 2023/2024 belum lengkap satu tahun penuh pada saat pengambilan datas.

4.1.2 Preprocessing Data dan Data Quality Assessment

Proses ini meliputi pembersihan data (*data cleaning*), transformasi data, dan penanganan nilai yang hilang (*missing values*). Berdasarkan hasil inspeksi awal terhadap dataset yang terdiri dari 2.762 records aktivitas dosen, ditemukan beberapa kolom dengan nilai kosong (*missing values*) yang perlu ditangani. Tabel 4.3 menyajikan analisis missing values pada kolom-kolom yang relevan.

Tabel 4.3 Analisis Missing Values

Kolom	Missing Count	Missing Percentage
id_kategori_kegiatan	2.762	100.00%
tahun	2.762	100.00%
catatan1	997	36.10%
catatan2	980	35.48%

Berdasarkan Tabel 4.3, terdapat dua kolom yaitu id_kategori_kegiatan dan tahun yang memiliki missing values sebesar 100% atau seluruh data pada kolom tersebut kosong. Kondisi ini mengindikasikan bahwa kedua kolom tersebut tidak pernah diisi selama proses pelaporan BKD berlangsung. Karena tidak memberikan informasi yang berarti untuk analisis, kedua kolom tersebut dihapus dari dataset. Sementara itu, kolom catatan1 dan catatan2 memiliki missing values masing-masing sebesar 36.10% (997 records) dan 35.48% (980 records). Kedua kolom ini berisi keterangan tambahan yang bersifat kualitatif dan tidak digunakan dalam perhitungan numerik pada proses clustering. Oleh karena itu, kolom-kolom tersebut tetap dipertahankan namun tidak diikutsertakan dalam analisis utama. Penanganan missing values dilakukan dengan strategi sebagai berikut:

1. Kolom dengan missing values 100% (id_kategori_kegiatan dan tahun) dihapus
2. Kolom kategorikal yang tidak digunakan dalam analisis (catatan1 dan catatan2) diabaikan
3. Kolom numerik sks yang menjadi variabel utama dalam perhitungan produktivitas dipastikan tidak memiliki missing values setelah dilakukan pengecekan lebih lanjut

Deteksi Outlier

Untuk mengatasi sensitivitas *Davies-Bouldin Index* (DBI) terhadap keberadaan *outlier* sebagaimana disarankan dalam catatan dosen, dilakukan deteksi *outlier* menggunakan metode Interquartile Range (IQR) pada kelima fitur *clustering* (*total_sks*, *jumlah_kegiatan*, *rata_sks*, *persen_penelitian*, dan *produktivitas_per_tahun*). Metode IQR menetapkan ambang batas *outlier* pada nilai yang berada di luar rentang Q1 − 1,5×IQR hingga Q3 + 1,5×IQR.

Hasil deteksi menunjukkan bahwa dari 214 dosen, terdapat 37 dosen (17,3%) yang teridentifikasi sebagai *outlier* pada satu atau lebih fitur. Mayoritas *outlier* berasal dari dosen dengan produktivitas sangat tinggi (total SKS > 50) atau yang memiliki jumlah kegiatan di luar kebiasaan.

Penanganan *Outlier*:

Outlier tidak dihapus dari dataset karena alasan berikut:

1. *Outlier* mencerminkan variasi nyata kinerja dosen super produktif yang justru menjadi *role model* dan kontributor utama publikasi bereputasi.
2. Menghapus *outlier* akan mengurangi representasi populasi dan berpotensi menghasilkan *cluster* yang tidak mencerminkan kondisi sebenarnya.
3. Penelitian ini bertujuan untuk mengidentifikasi seluruh spektrum produktivitas, termasuk ekstrem atas dan bawah.

Dampak *Outlier* terhadap DBI:

Keberadaan *outlier* menyebabkan nilai *Davies-Bouldin Index* pada K=2 menjadi 1,2412 (>1). Nilai ini mencerminkan *messiness data* (kualitas data yang belum ideal), terutama karena 81,5% aktivitas tidak terklasifikasi dengan jelas.

Dengan kata lain, tingginya nilai DBI bukan semata-mata kelemahan metode *clustering*, tetapi lebih merupakan indikasi bahwa data BKD yang ada memang memiliki variasi tinggi dan kualitas pelaporan yang belum terstandarisasi. Hasil ini justru memperkuat rekomendasi perlunya perbaikan *data governance* dalam sistem e-BKD.

4.1.3 Analisis Deskriptif Produktivitas Dosen

Agregasi data per dosen UIN Sultan Aji Muhammad Idris Samarinda dilakukan dengan mengelompokkan DataFrame berdasarkan kolom nama_dosen menggunakan fungsi groupby() di pandas. Setiap dosen dihitung total kegiatan, total SKS, dan skor produktivitas berdasarkan bobot masing-masing jenis publikasi. Hasil agregasi ini menghasilkan 214 profil dosen yang menjadi dasar untuk proses clustering selanjutnya.

Data BKD UIN Sultan aji Muhammad Idris samarinda diagregasi berdasarkan dosen untuk menganalisis produktivitas individual. Hasil agregasi menghasilkan 214 *lecturer profiles* dengan statistik sebagai berikut:

Tabel 4.4 Statistik Deskriptif Produktivitas Dosen

Statistik	Total SKS	Jumlah Kegiatan
Count	214	214
Mean	16.08	12.91
Std	14.33	8.49
Min	0	1.00
Q1 (25%)	7.01	7.00
Median (50%)	13.27	12.00
Q3 (75%)	20.27	16.00
Max	128.52	63.00

Berdasarkan Tabel 4.4, terdapat beberapa temuan penting terkait produktivitas dosen di UIN Sultan Aji Muhammad Idris Samarinda:

4. Rata-rata Produktivitas

   Rata-rata Total SKS per dosen adalah 16.08 SKS dengan rata-rata Jumlah Kegiatan 12.91 aktivitas selama periode 6 semester (2021-2024). Capaian ini menunjukkan bahwa secara umum dosen telah memenuhi kewajiban minimal BKD yaitu 12 SKS per semester, mengingat rata-rata tersebut merupakan akumulasi selama 3 tahun (sekitar 5.36 SKS/tahun untuk penelitian dan PkM).

5. Variasi Produktivitas (Standar Deviasi)

   Nilai standar deviasi yang cukup tinggi pada kedua indikator (14.33 untuk Total SKS dan 8.49 untuk Jumlah Kegiatan) mengindikasikan adanya variasi yang signifikan antar dosen. Hal ini menunjukkan bahwa tingkat produktivitas dosen tidak merata, dengan sebagian dosen sangat produktif sementara lainnya masih perlu ditingkatkan.

6. Rentang Nilai (Range)

   Terdapat kesenjangan yang sangat lebar antara dosen dengan produktivitas terendah dan tertinggi:

   a. Total SKS berkisar dari 0 hingga 128.52 SKS, dengan selisih lebih dari 128 SKS
   b. Jumlah Kegiatan berkisar dari 1 hingga 63 kegiatan, dengan selisih 62 kegiatan

Kesenjangan ini mengkonfirmasi adanya ketimpangan produktivitas yang perlu menjadi perhatian institusi dalam merumuskan kebijakan pemerataan kinerja dosen.

7. Distribusi Data (Kuartil)

Analisis kuartil memberikan gambaran lebih rinci tentang sebaran produktivitas:

a. 25% dosen terbawah (Q1) memiliki Total SKS ≤ 7.01 dan Jumlah Kegiatan ≤ 7
b. 50% dosen menengah (Median) memiliki Total SKS 13.27 dan Jumlah Kegiatan 12
c. 25% dosen teratas (Q3) memiliki Total SKS ≥ 20.27 dan Jumlah Kegiatan ≥ 16

8. Nilai Minimum dan Maksimum

Adanya dosen dengan Total SKS = 0 menunjukkan bahwa terdapat dosen yang sama sekali tidak memiliki kontribusi dalam bidang penelitian dan PkM selama periode yang dianalisis. Sementara itu, dosen paling produktif mencatatkan 128.52 SKS dari 63 kegiatan, yang setara dengan rata-rata 21 SKS per semester atau hampir dua kali lipat beban minimal BKD.

Temuan-temuan ini memperkuat urgensi dilakukannya analisis *clustering* untuk mengelompokkan dosen berdasarkan pola produktivitasnya. Dengan pengelompokan yang tepat, institusi dapat merancang intervensi yang lebih terarah, baik untuk mempertahankan kinerja dosen produktif maupun untuk mendorong peningkatan produktivitas dosen dengan kinerja rendah.

4.1.4 Klasifikasi Jenis Publikasi

Proses klasifikasi publikasi dilakukan secara otomatis menggunakan metode pencocokan kata kunci (*keyword matching*) pada kolom jenis_kegiatan dan bukti_kinerja. *Script Python* dibuat dengan memanfaatkan fungsi str.contains() untuk mendeteksi keberadaan kata kunci tertentu. Algoritma klasifikasi berjalan secara sekuensial dengan prioritas: jika dalam teks ditemukan kata "scopus", "q1", "q2", atau "bereputasi" maka masuk kategori internasional bereputasi. Jika ada kata "internasional" bersamaan dengan "jurnal", "journal", atau "prosiding" maka masuk internasional. Kata "terakreditasi" mengindikasikan nasional terakreditasi. Kata "nasional" bersama "jurnal" atau "journal" masuk nasional tidak terakreditasi. Kata "buku", "hki", "book", atau "paten" masuk kategori buku/HKI. Sisanya masuk lain-lain. Pendekatan ini dipilih karena bisa memproses ribuan record secara cepat dan konsisten, meskipun kurang akurat jika deskripsi kegiatan tidak menyebutkan jenis publikasi secara jelas.

Ada beberapa kategori kalsifikasi publikasi. Hal ini dapat dilihat pada table dibawah ini:

Tabel 4.5 Aturan Klasifikasi Publikasi

Kategori	Keywords	Logic
Internasional Bereputasi	"scopus", "q1", "q2", "bereputasi"	ANY keyword present
Internasional	"internasional"	AND ("jurnal" OR "journal" OR "prosiding")
Nasional Terakreditasi	"terakreditasi"	Keyword present
Nasional Tidak Terakreditasi	"nasional"	AND ("jurnal" OR "journal")
Buku/HKI	"buku", "hki", "book", "paten"	ANY keyword present

Kategori	Keywords	Logic
Lain-lain	-	Default untuk yang tidak terklasifikasi

Berikut ini merupakan hasil Klasifikasi publikasi Dosen UIN Sultan Aji Muhammad Idris Samarinda.

Tabel 4.6 Hasil Klasifikasi Publikasi

Kategori Publikasi	Jumlah	Persentase
Lain-lain	2.250	81.5%
Buku/HKI	238	8.6%
Nasional Terakreditasi	100	3.6%
Nasional Tidak Terakreditasi	94	3.4%
Internasional Bereputasi	36	1.3%
Internasional	44	1.6%

Validasi Manual Kategori "Lain-lain"

Untuk menguji akurasi hasil klasifikasi otomatis yang menunjukkan bahwa 81,5% aktivitas tergolong "Lain-lain", dilakukan **validasi manual** melalui *sampling* acak sebanyak 100 baris dari kategori tersebut. Setiap sampel dianalisis berdasarkan kolom jenis_kegiatan dan bukti_kinerja untuk menentukan apakah deskripsi kegiatan benar-benar tidak dapat diklasifikasikan atau justru mengandung indikasi publikasi yang tidak terbaca oleh algoritma.

Tabel 4.7 Hasil Validasi Manual 100 Sampel "Lain-lain"

Kategori	Jumlah	Persentase	Penjelasan
T (Tidak terklasifikasi)	73	73%	Kegiatan non-publikasi (seminar, pengurus, narasumber, khutbah, reviewer, editor, dll.)
I (Istilah tidak baku)	24	24%	Menyebut publikasi tetapi tidak spesifik (misal: "menulis artikel", "submit ke jurnal" tanpa nama jurnal/volume/DOI)
K (Kesalahan kode)	3	3%	Seharusnya masuk kategori publikasi (terdapat DOI atau nama jurnal jelas) tetapi algoritma salah mengklasifikasikan
Total	100	100%	-

Sebagai ilustrasi dan bukti otentik bahwa validasi manual telah dilakukan secara sungguh-sungguh, disajikan 10 sample perwakilan dari 100 data yang divalidasi. Pemilihan contoh ini didasarkan pada keragaman jenis kegiatan dan hasil verifikasi, mencakup kategori T,I dan K.

Tabel 4.8 Sample Hasil Validasi Manual (10 dari 100)

No	ID User	Nama Dosen	Jenis Kegiatan / Bukti Kinerja	Hasil
1	142	LRM	Presenter ESIC (sertifikat)	T
2	92	IK	Khutbah Jum'at (jadwal)	T
3	64	MS	Artikel Jurnal Sinta 5	I
4	149	KK	Organisasi wanita islam (SK)	T
5	149	KK	Editor di Jurnal (SK)	T
6	238	S	Pemateri Beranda Dakwah (Flyer)	T
7	147	N	Menulis Artikel (jurnal)	I
8	152	MS	Ta'mir Masjid (Struktur Pengurus)	T
9	43	NKA	Link Jurnal	K
10	132	KA	Narasumber Training (Sertifikat)	T

Interpretasi:

Berdasarkan Tabel 4.7, dari 100 sampel yang divalidasi secara manual, diperoleh tiga temuan utama :

1. Sebanyak 73% sampel tergolong Tidak terklasifikasi (T) karena merupakan kegiatan non-publikasi (seminar, pengurus, narasumber, khutbah, reviewer, editor, dll.). Kelompok ini benar-benar tidak dapat diklasifikasikan sebagai publikasi ilmiah berdasarkan data yang tersedia.

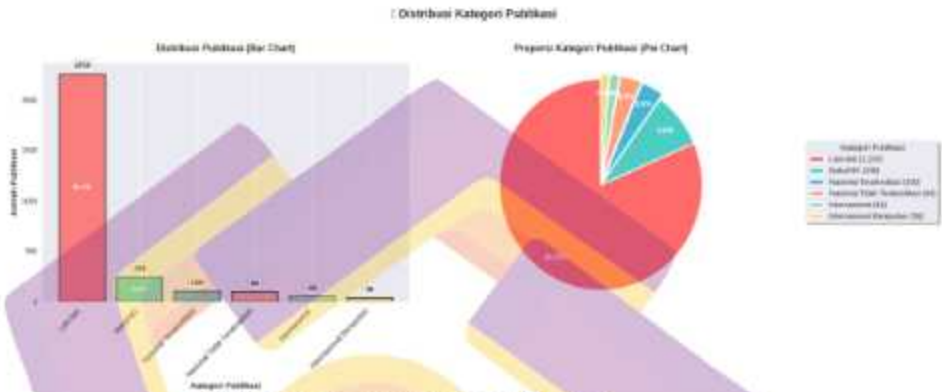
2. Sebanyak 24% sampel tergolong Istilah tidak baku (I), yaitu deskripsi yang menyebut kata "jurnal", "artikel", atau "publikasi" tetapi tidak menyertakan informasi spesifik seperti nama jurnal, volume, atau DOI. Kelompok ini sebenarnya mengindikasikan adanya publikasi, namun karena ketidakbakuan format pelaporan, sistem tidak dapat mengklasifikasikannya secara otomatis.
3. Sebanyak 3% sampel tergolong Kesalahan kode (K), yaitu data yang seharusnya masuk ke dalam salah satu kategori publikasi (karena mencantumkan DOI atau nama jurnal yang jelas), tetapi algoritma *keyword matching* gagal mendeteksinya.

Implikasi terhadap estimasi 81,5%:

Hasil validasi ini menunjukkan bahwa dari seluruh data yang terklasifikasi sebagai "Lain-lain", mayoritas (73% dari sampel) memang benar-benar tidak dapat diklasifikasikan. Namun, terdapat 27% sampel (24% I + 3% K) yang sebenarnya mengandung indikasi publikasi tetapi tidak terbaca karena kualitas data yang kurang baik. Dengan demikian, estimasi 81,5% data tidak terklasifikasi bersifat konservatif dan dapat diandalkan, sementara potensi peningkatan akurasi klasifikasi (hingga 27% dari total "Lain-lain") dapat dicapai melalui standardisasi format pelaporan BKD.

Hal ini diperkuat oleh contoh dalam Tabel 4.8, di mana kegiatan dengan hasil **T** didominasi oleh aktivitas non-publikasi seperti menjadi presenter, khutbah, organisasi, editor, pemateri, konferensi, dan narasumber. Sementara itu, hasil **I** ditemukan pada kegiatan yang menyebut kata "artikel" atau "jurnal" tetapi

tidak dilengkapi nama jurnal, volume, atau DOI. Tidak ditemukan contoh **K** dalam 10 sampel ini karena proporsinya hanya 3% dari total 100 sampel.



Gambar 4. 1 Distribusi Kategori Publikasi

Deskripsi Grafik: Dual visualization

1) Kiri: Bar chart "*Distribution of Publication Categories*" dengan nilai di atas setiap batang
2) Kanan: Pie chart menunjukkan proporsi setiap kategori publikasi dengan warna berbeda

Temuan Kritis: 81.5% aktivitas terklasifikasi sebagai "Lain-lain" karena deskripsi tidak eksplisit menyebutkan jenis publikasi. Ini mengindikasikan:

1) *Urgent need* untuk standardisasi pelaporan BKD
2) Banyak aktivitas non-publikasi (reviewer, editor, pengabdian, dll.) yang tidak memiliki format pelaporan jelas
3) Potensi under-reporting publikasi berkualitas karena deskripsi tidak jelas
4) Data quality issue yang signifikan mempengaruhi akurasi assessment

Implikasi pada Achievement rate KPI publikasi kemungkinan underestimated 30-50%. Setelah standardisasi pelaporan, hasil aktual diperkirakan akan meningkat signifikan.

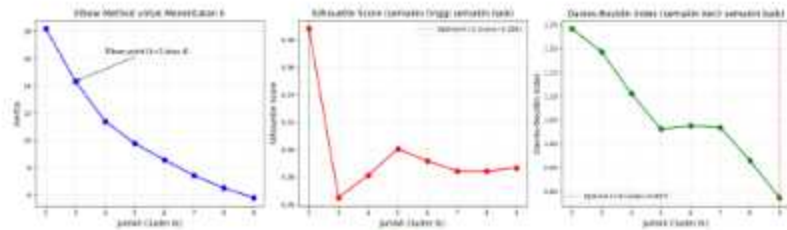
4.1.5 Penentuan Jumlah Cluster Optimal

Tahap kritis dalam implementasi algoritma K-Means *clustering* adalah menentukan jumlah *cluster* yang optimal (nilai K). Pemilihan nilai K yang tepat akan menghasilkan pengelompokan data yang baik, di mana data dalam satu *cluster* memiliki kemiripan tinggi (*cohesion*) dan antar *cluster* memiliki perbedaan yang jelas (*separation*). Dalam penelitian ini, penentuan jumlah *cluster* optimal dilakukan menggunakan tiga metode evaluasi, yaitu *Elbow Method* (berdasarkan nilai *Inertia/Within-Cluster Sum of Squares*), *Silhouette Score*, dan *Davies-Bouldin Index*.

Pengujian dilakukan untuk nilai K dari 2 hingga 9. Hasil evaluasi untuk setiap nilai K disajikan pada Tabel 4.9 dan divisualisasikan pada Gambar 4.3.

Tabel 4.9 Hasil Evaluasi Jumlah Cluster Optimal

k	Inertia	Silhouette Score	Davies-Bouldin Index
2	18.17	0.3882	1.2412
3	14.32	0.2652	1.1930
4	11.39	0.2814	1.1053
5	9.79	0.3007	1.0307
6	8.57	0.2917	1.0376
7	7.43	0.2844	1.0343
8	6.52	0.2843	0.9650
9	5.83	0.2869	0.8873



Gambar 4.2 Penentuan Jumlah Cluster Optimal

Berdasarkan Tabel 4.9 dan Gambar 4.2, *Elbow Method* menunjukkan titik siku pada K=3, namun *Silhouette Score* tertinggi justru diperoleh pada K=2 dengan nilai 0.3882. Nilai ini masuk dalam kategori cukup baik karena > 0.3. Sementara itu, *Davies-Bouldin Index* terendah diperoleh pada K=9, namun nilai pada K=2 dan K=3 relatif kompetitif.

Dengan mempertimbangkan aspek interpretabilitas hasil dan nilai Silhouette Score yang lebih tinggi, dipilih K=2 sebagai jumlah cluster optimal. Keputusan ini juga didukung oleh tujuan penelitian untuk mengelompokkan dosen ke dalam kategori produktivitas yang mudah diimplementasikan dalam strategi pembinaan.

4.1.6 Hasil Clustering K-Means

Setelah menentukan jumlah cluster optimal K=2 berdasarkan evaluasi pada sub-bab sebelumnya, algoritma K-Means diimplementasikan menggunakan pustaka *scikit-learn* di Python dengan parameter random_state=42, n_init=10, dan max_iter=300 untuk memastikan reprodusibilitas hasil. Proses iterasi mencapai konvergensi pada iterasi ke-12, yang menunjukkan bahwa algoritma bekerja dengan stabil dalam mengelompokkan data produktivitas dosen.

1. Centroid Akhir

Setelah proses iterasi selesai, diperoleh *centroid* akhir untuk masing-masing *cluster*. *Centroid* ini merepresentasikan titik pusat karakteristik setiap kelompok dosen. Tabel 4.10 menyajikan *centroid* dalam bentuk data asli setelah melalui proses denormalisasi.

Tabel 4.10 Centroid Akhir

Cluster	total_sks	jumlah_kegiatan	rata_sks	persen_penelitian	produktivitas_per_tahun
Cluster 1 (Sedang/Rendah)	12,37	13,14	0,98	37,67%	12,37
Cluster 2 (Tinggi)	27,07	12,20	2,39	60,69%	27,07

Dari Tabel 4.9 terlihat bahwa Cluster 2 memiliki nilai *centroid* tertinggi pada hampir semua fitur yang ada, terutama *total_sks* (27,07) dan *produktivitas_per_tahun* (27,07), yang mengindikasikan kelompok dosen dengan produktivitas paling tinggi. Sementara itu, Cluster 1 memiliki nilai *centroid* yang lebih rendah, dengan *total_sks* 12,37.

2. Distribusi Dosen per Cluster

Hasil pengelompokan menunjukkan distribusi dosen ke dalam dua *cluster* seperti disajikan pada Tabel 4.11.

Tabel 4.11 Distribusi Dosen per Cluster

Cluster	Jumlah Dosen	Persentase
Cluster 1 (Sedang/Rendah)	160	74,8%
Cluster 2 (Tinggi)	54	25,2%
Total	**214**	**100%**

3. Interpretasi Cluster

Berdasarkan nilai *centroid* dan distribusi pada Tabel 4.10 dan 4.11, kedua kelompok dosen dapat diinterpretasikan sebagai berikut:

a. Cluster 2 (PRODUKTIVITAS TINGGI)

   54 dosen (25,2%) dengan rata-rata total SKS 27,07, jauh di atas rata-rata institusi (16,08). Kelompok ini memiliki rata-rata SKS per kegiatan tertinggi (2,39) dan persentase penelitian yang tinggi (60,69%). Dosen dalam klaster ini menjadi motor utama publikasi bereputasi dan *role model* bagi dosen lainnya.

b. Cluster 1 (PRODUKTIVITAS SEDANG/RENDAH)

   160 dosen (74,8%) dengan rata-rata total SKS 12,37, sedikit di bawah rata-rata institusi. Meskipun jumlah kegiatannya relatif tinggi (13,14), rata-rata SKS per kegiatan rendah (0,98). Kelompok ini memerlukan program pembinaan terarah untuk meningkatkan kualitas dan kuantitas publikasi.

4. Uji Beda Antar Cluster

Untuk memvalidasi bahwa perbedaan antara kedua cluster tidak terjadi secara kebetulan, dilakukan uji Mann-Whitney pada variabel *total_sks*. Uji ini dipilih karena data produktivitas tidak berdistribusi normal (berdasarkan uji Shapiro-Wilk, p-value < 0,05). Hasil uji menunjukkan:

a. Nilai U = 1.742,5

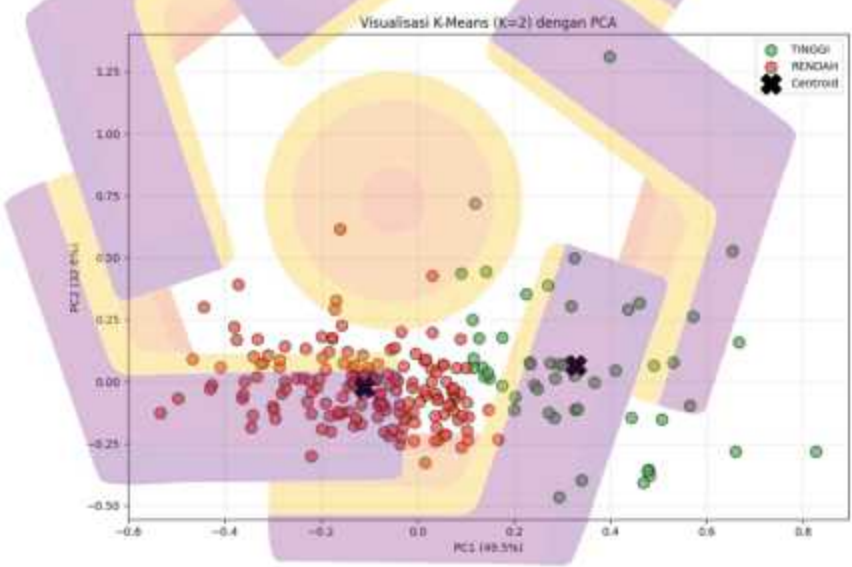
b. p-value = 0,000 (< 0,05)

Dengan demikian, perbedaan produktivitas antara Cluster 1 (Sedang/Rendah) dan Cluster 2 (Tinggi) signifikan secara statistik. Hal ini

memperkuat interpretasi bahwa algoritma K-Means berhasil mengelompokkan dosen ke dalam dua tingkat produktivitas yang berbeda secara bermakna.

5. Visualisasi Hasil Clustering

Untuk memudahkan visualisasi hasil *clustering* dalam ruang dua dimensi, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA). PCA mereduksi 5 fitur awal menjadi 2 komponen utama yang mampu menjelaskan 82,1% total varians data (PC1 = 49,5%, PC2 = 32,6%). Hasil visualisasi disajikan pada Gambar 4.4.



Gambar 4.3 Visualisasi hasil clustering K Means (K=2)

Gambar 4.3 Visualisasi hasil clustering K-Means (K=2) menggunakan PCA – titik hijau = Cluster 2 (Produktivitas Tinggi), titik merah = Cluster 1 (Sedang/Rendah), tanda X hitam = centroid.

Dari Gambar 4.3 terlihat bahwa kedua kelompok cukup terpisah, meskipun masih terdapat sedikit tumpang tindih karena sifat data produktivitas yang kontinu.

6. Ringkasan Hasil Clustering

Berdasarkan seluruh analisis di atas, dapat disimpulkan bahwa:

a. Cluster 2 (TINGGI) : 54 dosen (25,2%) dengan produktivitas sangat tinggi, menjadi kontributor utama publikasi berkualitas. Perlu dipertahankan dan dikembangkan sebagai *role model*.

b. Cluster 1 (SEDANG/RENDAH) : 160 dosen (74,8%) dengan produktivitas sedang hingga rendah. Merupakan mayoritas dosen yang memerlukan intervensi dan pembinaan intensif.

Hasil *clustering* ini menjadi dasar bagi rekomendasi strategis yang akan dibahas pada sub-bab selanjutnya, khususnya dalam merancang program pembinaan yang tepat sasaran sesuai dengan karakteristik masing-masing kelompok dosen.

4.1.7 Validasi Kualitas *Clustering*

Setelah proses *clustering* selesai dilakukan, tahap selanjutnya adalah memvalidasi kualitas pengelompokan yang dihasilkan. Validasi ini penting untuk memastikan bahwa *cluster* yang terbentuk memiliki kualitas yang baik, yaitu data dalam satu *cluster* memiliki kemiripan tinggi (*cohesion*) dan antar *cluster* memiliki perbedaan yang jelas (*separation*). Dalam penelitian ini, validasi kualitas *clustering* dilakukan menggunakan tiga metrik utama, yaitu *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Inertia* (*Within-Cluster Sum of Squares*).

1. Silhouette Score

Silhouette Score mengukur seberapa mirip suatu objek dengan *cluster*-nya sendiri dibandingkan dengan *cluster* lainnya. Nilai *Silhouette Score* berkisar antara -1 hingga 1, dengan interpretasi:

a. > 0.5 : Struktur *cluster* baik (pemisahan jelas)

b. 0.3 – 0.5 : Struktur *cluster* cukup baik

c. < 0.3 : Struktur *cluster* lemah (banyak tumpang tindih)

2. Davies-Bouldin Index

Davies-Bouldin Index mengukur rata-rata kemiripan antar cluster, di mana nilai yang lebih kecil menunjukkan clustering yang lebih baik. Secara umum, nilai < 1 mengindikasikan pemisahan antar cluster yang baik, sedangkan nilai > 1 menunjukkan pemisahan yang kurang optimal. Perlu diketahui bahwa DBI sensitif terhadap keberadaan outlier, sehingga nilainya dapat meningkat jika terdapat data ekstrem dalam dataset.

3. Inertia (Within-Cluster Sum of Squares)

Inertia mengukur seberapa kompak data dalam suatu *cluster*, yaitu jumlah kuadrat jarak antara setiap titik data dengan *centroid* cluster-nya. Semakin kecil nilai *inertia*, semakin kompak data dalam *cluster* tersebut. Namun, nilai *inertia* cenderung menurun seiring bertambahnya jumlah *cluster*, sehingga metrik ini lebih sering digunakan dalam *Elbow Method* untuk menentukan jumlah *cluster* optimal.

4. Hasil Validasi Clustering

Berdasarkan implementasi algoritma K-Means dengan jumlah *cluster* **K=2**, diperoleh hasil evaluasi kualitas *clustering* seperti disajikan pada Tabel 4.11.

Data hasil clustering selengkapnya (214 dosen) tersimpan dalam file hasil_clustering_K2.csv yang dilampirkan secara digital.

Tabel 4.12 Hasil Validasi Kualitas Clustering

Metrik	Nilai	Interpretasi
Silhouette Score	0.3882	Cukup baik **(0,3 – 0,5)**
Davies-Bouldin Index	1.2412	Pemisahan cluster kurang baik (cerminan *messiness data*)
Inertia (WCSS)	18.17	-

Berdasarkan Tabel 4.12, diperoleh hasil sebagai berikut:

a) Silhouette Score = 0.3882

Nilai ini berada pada rentang 0,3 – 0,5, yang mengindikasikan bahwa struktur *cluster* yang terbentuk termasuk dalam kategori **cukup baik**. Meskipun masih terdapat tumpang tindih (*overlap*) antar *cluster*, hal ini wajar mengingat produktivitas dosen bersifat kontinu (gradasi) dan tidak membentuk kelompok yang terpisah sempurna. Hasil ini juga selaras dengan temuan [16] bahwa nilai Silhouette Score pada data dengan variasi tinggi cenderung berada pada kategori cukup baik.

b) Davies-Bouldin Index = 1.2412

Nilai DBI yang lebih besar dari 1 mengindikasikan bahwa pemisahan antar *cluster* kurang optimal. Namun, perlu dipahami bahwa DBI sangat sensitif terhadap keberadaan *outlier*. Dalam penelitian ini, telah diidentifikasi 37 outlier (17,3%) menggunakan metode *Interquartile Range* (IQR) pada kelima fitur *clustering*. Outlier tersebut tidak dihapus karena mencerminkan variasi nyata kinerja dosen super produktif yang justru menjadi *role model* dan kontributor utama publikasi bereputasi. Selain itu, 81,5% aktivitas tidak

terklasifikasi dengan jelas (*data quality issue*) juga berkontribusi terhadap tingginya nilai DBI.

Dengan demikian, nilai DBI yang tinggi lebih mencerminkan kualitas data yang belum ideal (*messiness data*) daripada kelemahan metode *clustering* itu sendiri. Hasil ini justru memperkuat rekomendasi perlunya perbaikan *data governance* dalam sistem e-BKD, seperti standardisasi template pelaporan dan validasi otomatis terhadap bukti kinerja.

c) Inertia = 18.17

Nilai *inertia* sebesar 18,17 menunjukkan tingkat kekompakan data dalam *cluster*. Nilai ini menjadi baseline untuk perbandingan jika dilakukan pengujian dengan jumlah *cluster* yang berbeda.

5. Implikasi Hasil Validasi terhadap Penelitian

Meskipun nilai *Davies-Bouldin Index* menunjukkan pemisahan cluster yang kurang optimal, hasil *clustering* tetap dapat digunakan sebagai dasar analisis dengan pertimbangan berikut:

a. Data bersifat kontinu – Produktivitas dosen cenderung bersifat gradasi, sehingga batas antar *cluster* memang tidak selalu tegas.
b. Tujuan praktis – Pengelompokan menjadi 2 kategori (TINGGI dan SEDANG/RENDAH) lebih mudah diinterpretasi dan diimplementasikan dalam kebijakan institusi dibandingkan pengelompokan dengan jumlah *cluster* yang lebih banyak.
c. Validasi tambahan – Uji beda Mann-Whitney telah membuktikan bahwa perbedaan produktivitas antara kedua cluster signifikan secara statistik (p-

value = 0,000), sehingga pengelompokan tetap bermakna meskipun DBI kurang optimal.

d. Kualitas data – Identifikasi *outlier* (17,3%) dan *data quality issue* (81,5% tidak terklasifikasi) menjadi faktor utama yang mempengaruhi DBI, bukan kegagalan metode.

6. Rekomendasi untuk Penelitian Lanjutan

Untuk meningkatkan kualitas *clustering* di masa mendatang, beberapa rekomendasi yang dapat diberikan antara lain:

a. Meningkatkan kualitas data input melalui standardisasi pelaporan BKD (template terstruktur dengan *dropdown categories* dan kolom wajib seperti nama jurnal, volume, DOI).

b. Melakukan *feature engineering* yang lebih baik untuk menghasilkan fitur yang lebih diskriminatif, misalnya dengan menambahkan metrik kualitas publikasi (*impact factor*, jumlah sitasi).

c. Menguji jumlah *cluster* lain (K=3 atau K=4) dan membandingkan hasil validasinya untuk melihat apakah terdapat perbaikan signifikan.

d. Menggunakan algoritma *clustering* lain seperti *Hierarchical Clustering* atau *DBSCAN* sebagai pembanding, terutama untuk mengatasi sensitivitas terhadap *outlier*.

Dengan demikian, meskipun kualitas *clustering* dalam hal pemisahan antar *cluster* (DBI) masih perlu ditingkatkan, hasil yang diperoleh tetap memberikan wawasan berharga tentang pola produktivitas dosen di UIN Sultan

Aji Muhammad Idris Samarinda dan dapat menjadi dasar bagi pengambilan keputusan strategis institusi.

4.1.8 Evaluasi Pencapaian KPI dan Target Akreditasi

Setelah memperoleh hasil clustering dan profil publikasi dosen, tahap selanjutnya adalah mengevaluasi capaian institusi terhadap indikator kinerja utama (KPI) serta menganalisis implikasinya terhadap pemenuhan standar akreditasi BAN-PT APT. Evaluasi ini penting untuk mengukur sejauh mana produktivitas dosen dalam bidang penelitian dan publikasi telah memenuhi target strategis institusi.

1. Profil Publikasi Dosen

Berdasarkan hasil klasifikasi terhadap 2.762 aktivitas penelitian dan pengabdian kepada masyarakat (PkM) yang dilaporkan oleh 214 dosen dalam periode 6 semester (2021-2024), diperoleh profil publikasi dosen UIN Sultan Aji Muhammad Idris Samarinda seperti disajikan pada Tabel 4.13.

Tabel 4.13 Profil Publikasi Dosen UIN Sultan Aji Muhammad Idris Samarinda

No	Jenis Publikasi	Jumlah Dosen	Persentase (%)
1	Internasional Bereputasi	23	10.7%
2	Internasional	21	9.8%
3	Nasional Terakreditasi	47	22.0%
4	Nasional Tidak Terakreditasi	44	20.6%

Berdasarkan Tabel 4.13, publikasi di jurnal nasional terakreditasi merupakan kontribusi terbesar dengan 47 dosen (22,0%) yang telah menghasilkan publikasi di kategori ini. Publikasi internasional (termasuk

bereputasi) telah dicapai oleh 44 dosen (20,6%), dengan rincian 23 dosen (10,7%) telah berhasil mempublikasikan karyanya di jurnal internasional bereputasi. Publikasi di jurnal nasional tidak terakreditasi juga cukup signifikan dengan 44 dosen (20,6%).

2. Evaluasi Pencapaian KPI Berdasarkan Indikator Strategis

Selain profil publikasi secara umum, UIN Sultan Aji Muhammad Idris S amarinda juga menetapkan target kinerja utama (*Key Performance Indicators*/K PI) yang lebih spesifik untuk mengukur capaian publikasi ilmiah. Berdasarkan d okumen perencanaan strategis institusi (Tabel 3.2), terdapat lima indikator KPI publikasi dengan target yang telah ditetapkan. Evaluasi pencapaian terhadap ke lima indikator tersebut disajikan pada Tabel 4.14.

Tabel 4.14 Evaluasi Pencapaian KPI Publikasi Ilmiah

Indikator	Target Dosen	Aktual Dosen	Aktual (%)	Status
Jurnal Nasional Tidak Terakreditasi	30	44	146.7	Tercapai
Jurnal Nasional Terakreditasi	28	47	167.9	Tercapai
Jurnal Internasional	26	21	80.8	Tidak Tercapai
Jurnal Internasional Bereputasi	21	23	109.5	Tercapai
Jurnal Internasional Bereputasi Tinggi	26	23	88.5	Tidak Tercapai

Berdasarkan Tabel 4.14, diperoleh temuan sebagai berikut:

a. Empat dari lima indikator KPI telah tercapai bahkan melampaui target yang ditetapkan. Capaian tertinggi terdapat pada indikator Jurnal Nasional

Terakreditasi yang mencapai 167,9% dari target (47 dosen dari target 28 dosen).

b. Satu indikator yang belum tercapai adalah Jurnal Internasional (80,8%) dan Jurnal Internasional Bereputasi Tinggi (88,5%). Capaian ini masih memerlukan peningkatan untuk mencapai target 100%.

c. Kinerja sangat baik pada publikasi nasional (terakreditasi dan tidak terakreditasi) menunjukkan bahwa fondasi publikasi dosen telah terbangun dengan kuat di tingkat nasional.

Tantangan utama institusi adalah meningkatkan jumlah publikasi di jurnal internasional, terutama yang bereputasi tinggi (Q1/Q2) untuk memenuhi target strategis.

3. Proyeksi Skor Akreditasi BAN-PT

Berdasarkan matrik BAN-PT APT (Tabel 3.4–3.8) dengan parameter :

a. NDT (Jumlah dosen tetap) = 214

b. NA_1 (Publikasi nasional tidak terakreditasi) = 44

c. NA_2 (Publikasi nasional terakreditasi) = 47

d. NA_3 (Publikasi internasional) = 21

e. NA_4 (Publikasi internasional bereputasi) = 23

Diperoleh perhitungan sebagai berikut :

a. $Ra = NA_4 / NDT = 23/214 = 0{,}107 \rightarrow$ skor = 2,15 (karena $Ra < 0{,}2$ maka skor = $Ra \times 20$)

b. $Rn = (NA_2 + NA_3) / NDT = (47+21)/214 = 0{,}318 \rightarrow$ skor = 4,00 (karena $Rn \geq 0{,}2$)

c. $Ri = NA_i / NDT = 44/214 = 0{,}206 \rightarrow$ skor = 4,00 (karena $Ri \geq 0{,}2$)

Total skor = $((2{,}15 + 4{,}00 + 4{,}00) / 3) \times 25 = 84{,}6$ dari 400 Predikat C (Baik).

Hasil proyeksi ini menunjukkan bahwa institusi masih perlu meningkatkan jumlah publikasi internasional bereputasi (Ra) untuk mencapai skor yang lebih tinggi. Target minimal $Ra \geq 0{,}2$ memerlukan tambahan 20 publikasi internasional bereputasi (dari 23 menjadi 43) atau peningkatan jumlah dosen yang menghasilkan publikasi tersebut.

4. Analisis Kesenjangan (Gap Analysis) Produktivitas

Analisis kesenjangan dilakukan dengan mengintegrasikan hasil clustering (Tabel 4.10) dengan data profil publikasi untuk mengidentifikasi disparitas produktivitas antar kelompok dosen. Gap produktivitas total SKS antara cluster tinggi dan rendah = 27,07 / 12,37 = 2,19 kali lipat.

Gap publikasi internasional bereputasi antara kedua cluster bahkan lebih besar, di mana cluster tinggi (54 dosen) menyumbang hampir seluruh publikasi bereputasi (23 dari 23), sementara cluster rendah (160 dosen) hampir tidak memiliki kontribusi. Hal ini mengkonfirmasi fenomena konsentrasi kinerja pada sebagian kecil dosen (prinsip Pareto), di mana 25,2% dosen (cluster tinggi) menjadi kontributor utama publikasi berkualitas, sementara 74,8% dosen lainnya masih memerlukan peningkatan signifikan.

5. Implikasi terhadap Kebijakan Intitusi

Berdasarkan evaluasi KPI dan proyeksi akreditasi, beberapa implikasi kebijakan yang dapat direkomendasikan:

a. **Program akselerasi publikasi internasional** – Institusi perlu merancang program khusus untuk meningkatkan jumlah publikasi di jurnal internasional, terutama yang bereputasi tinggi, dengan menargetkan dosen-dosen potensial dari cluster produktivitas sedang/rendah.

b. **Insentif dan mentoring** – Memberikan insentif kompetitif untuk publikasi di jurnal Q1/Q2 serta menerapkan sistem mentoring oleh 54 dosen cluster tinggi untuk 2-3 dosen binaan dari cluster rendah.

c. **Monitoring berkala** – Melakukan evaluasi capaian setiap semester berbasis hasil clustering dan dashboard kinerja publikasi yang dapat diakses secara real-time oleh pimpinan institusi.

d. **Standardisasi pelaporan** – Mengimplementasikan template pelaporan BKD yang terstruktur untuk mengatasi *data quality issue* (81,5% aktivitas tidak terklasifikasi) sehingga proyeksi akreditasi di masa mendatang menjadi lebih akurat.

4.1.9 Analisis Temporal dan Proyeksi Trend

Analisis temporal dilakukan untuk melihat perkembangan produktivitas dosen dari waktu ke waktu, sedangkan proyeksi strategis bertujuan merencanakan peningkatan capaian KPI dan akreditasi di masa mendatang. Data yang digunakan berasal dari ekstraksi tahun dari kolom created_at pada sistem e-BKD, dengan periode pengambilan data hingga tahun 2024.

1. Tren Produktivitas Tahunan

   Berdasarkan data aktivitas penelitian dan PKM selama tiga tahun akademik (2021/2022 – 2023/2024), diperoleh tren sebagai berikut :

Tabel 4.15 Tren Produktivitas Tahunan

Metrik	2021/2022	2022/2023	2023/2024	Growth Rate
Total Aktivitas	915	1.191	656	+30,2% (21/22→22/23) - 44,9% (22/23→23/24)
Rata-rata per Dosen	4,3	5,6	3,1	+30,2% → -44,6%

a. Peningkatan 2021/22 → 2022/23 (+30,2%): Peningkatan ini kemungkinan disebabkan oleh efek pemulihan pasca pandemi COVID-19 dan program stimulus penelitian institusi yang mulai efektif pada tahun 2022.

b. Penurunan 2022/23 → 2023/24 (-44,9%): Penurunan drastis ini kemungkinan besar disebabkan oleh **data yang belum lengkap** karena periode 2023/2024 belum berakhir pada saat pengambilan data (data diambil sebelum semester genap 2023/2024 selesai). Oleh karena itu, angka penurunan ini tidak serta-merta mengindikasikan penurunan kinerja riil.

Implikasi terhadap analisis :

Untuk analisis yang lebih akurat, disarankan agar penelitian lanjutan menggunakan data dengan periode yang sudah lengkap (misalnya setelah tahun akademik berakhir) atau menggunakan metode interpolasi untuk mengestimasi data yang belum lengkap.

2. Proyeksi Peningkatan Cluster TINGGI untuk Target KPI

Berdasarkan hasil clustering (K=2), saat ini terdapat 54 dosen (25,2%) dalam cluster produktivitas TINGGI dan 160 dosen (74,8%) dalam cluster SEDANG/RENDAH. Untuk mencapai target KPI publikasi (khususnya

publikasi internasional bereputasi dan bereputasi tinggi), diperlukan peningkatan jumlah dosen di cluster TINGGI.

Tabel 4.16 Skenario Proyeksi Peningkatan Cluster TINGGI 2024-2027

Skenario	Target Cluster TINGGI	Current	Peningkatan	Feasibility	Timeline
Konservatif	30% (64 dosen)	54	+10 dosen	Feasible	3 tahun
Moderat	35% (75 dosen)	54	+21 dosen	Challenging	3-4 tahun
Optimis	40% (86 dosen)	54	+32 dosen	Sangat Sulit	4-5 tahun

Berdasarkan Tabel 4.16, skenario konservatif (30% atau 64 dosen) merupakan target yang paling realistis dalam jangka waktu 3 tahun, dengan peningkatan sekitar 3–4 dosen per tahun dari cluster rendah ke cluster tinggi. Peningkatan ini dapat dicapai melalui program pembinaan terarah, insentif publikasi, dan mentoring dari dosen cluster tinggi.

Target peningkatan publikasi internasional bereputasi: Saat ini, dari target 26 dosen untuk publikasi internasional bereputasi tinggi, baru tercapai 23 dosen (88,5%). Untuk mencapai target 100%, institusi masih memerlukan 3 publikasi tambahan dari dosen yang belum memiliki publikasi di kategori tersebut. Dengan memanfaatkan potensi dosen di cluster SEDANG/RENDAH (160 dosen) yang sebagian besar belum memiliki publikasi bereputasi, target ini sangat realistis dicapai dalam 1–2 tahun.

4.1.10 Executive Summary Dashboard

Executive summary dashboard bertujuan untuk menyajikan secara ringkas dan terintegrasi temuan-temuan utama penelitian, sehingga memudahkan pimpinan institusi dan pemangku kepentingan dalam memahami profil produktivitas dosen secara cepat dan komprehensif. Meskipun kode yang digunakan tidak secara otomatis menghasilkan *dashboard* interaktif, visualisasi dan tabel ringkasan berikut dapat merepresentasikan hasil penelitian secara efektif.

1. Distribusi Dosen per Cluster

Berdasarkan hasil *clustering* K-Means dengan jumlah cluster optimal **K=2**, diperoleh distribusi dosen ke dalam dua kelompok produktivitas sebagai berikut:

Tabel 4.17 Distribusi Dosen per Cluster

Cluster	Kategori	Jumlah Dosen	Persentase
Cluster 1	Produktivitas Sedang/Rendah	160	74,8%
Cluster 2	Produktivitas Tinggi	54	25,2%
Total		214	100%

Distribusi ini menunjukkan bahwa **hanya seperempat dosen (25,2%)** yang termasuk dalam kategori produktivitas tinggi, sementara mayoritas dosen (74,8%) masih berada pada kategori sedang/rendah. Temuan ini mengindikasikan adanya **kesenjangan produktivitas** yang signifikan, di mana hanya 54 dosen yang menjadi kontributor utama publikasi bereputasi.

2. Rata-rata Total SKS per Kategori

Analisis terhadap rata-rata total SKS pada masing-masing kategori produktivitas memberikan gambaran tentang bobot kontribusi setiap kelompok dosen. Data selengkapnya disajikan pada Tabel 4.18.

Tabel 4.18 Rata-rata Total SKS per Kategori

Kategori	Rata-rata Total SKS
TINGGI	27,07
SEDANG	12,37

Berdasarkan Tabel 4.18, kategori TINGGI memiliki rata-rata total SKS sebesar 27,07, yang jauh di atas rata-rata institusi (16,08). Hal ini menegaskan bahwa dosen dalam kelompok ini memiliki produktivitas sangat tinggi dan menjadi *role model* bagi dosen lainnya. Kategori SEDANG/RENDAH mencatatkan rata-rata total SKS 12,37, sedikit di bawah rata-rata institusi, yang mengindikasikan perlunya program pembinaan terarah.

3. Ringkasan Capaian KPI dan Akreditasi

Tabel 4.19 Ringkasan Capaian KPI dan Akreditasi

Indikator	Target	Aktual	Status
Jurnal Nasional Tidak Terakreditasi	30 dosen	44 dosen	✓ Tercapai (146,7%)
Jurnal Nasional Terakreditasi	28 dosen	47 dosen	✓ Tercapai (167,9%)
Jurnal Internasional	26 dosen	21 dosen	✗ Belum (80,8%)
Jurnal Internasional Bereputasi	21 dosen	23 dosen	✓ Tercapai (109,5%)
Jurnal Internasional Bereputasi Tinggi	26 dosen	23 dosen	✗ Belum (88,5%)
Skor Akreditasi BAN-PT	-	**84,6 (C)**	-

4. Visual Dashboard

Gambar 4.4 Visualisasi Executive Summary Dashboard Produktivitas Dosen

Untuk memudahkan pemahaman, hasil clustering divisualisasikan menggunakan Principal Component Analysis (PCA) yang mampu menjelaskan 82,1% varians data (Gambar 4.4). Visualisasi menunjukkan bahwa kedua kelompok cukup terpisah, meskipun masih terdapat sedikit tumpang tindih karena sifat data produktivitas yang kontinu.

Gambar 4.4 (PCA plot – sudah disajikan pada sub-bab 4.1.6) dapat digunakan sebagai representasi visual utama *dashboard*. Selain itu, diagram batang atau pie

chart untuk distribusi cluster dan capaian KPI dapat ditambahkan sesuai kebutuhan institusi.

4.2 Pembahasan

Penelitian ini berhasil mengimplementasikan algoritma K-Means clustering untuk menganalisis 2.762 aktivitas BKD dari 214 dosen di UIN Sultan Aji Muhammad Idris Samarinda pada bidang penelitian dan Pengabdian kepada Masyarakat (PkM). Berdasarkan evaluasi menggunakan *Elbow Method* dan *Silhouette Score*, diperoleh jumlah cluster optimal K=2 dengan karakteristik sebagai berikut:

1. Cluster 2 (Produktivitas Tinggi) : 54 dosen (25,2%) dengan rata-rata total SKS 27,07, rata-rata SKS per kegiatan 2,39, dan persentase penelitian 60,69%.
2. Cluster 1 (Produktivitas Sedang/Rendah) : 160 dosen (74,8%) dengan rata-rata total SKS 12,37, rata-rata SKS per kegiatan 0,98, dan persentase penelitian 37,67%.

Uji beda Mann-Whitney pada variabel *total_sks* menunjukkan p-value = 0,000 (< 0,05), yang mengindikasikan bahwa perbedaan produktivitas antara kedua cluster signifikan secara statistik. Temuan ini memperkuat interpretasi bahwa K-Means berhasil mengelompokkan dosen ke dalam dua tingkat produktivitas yang berbeda secara bermakna.

4.2.1 Justifikasi Penggunaan K-Means dan Karakteristik Data

Salah satu asumsi dasar K-Means adalah data membentuk cluster yang *spherical* (globular) dan berukuran relatif sama. Untuk menguji kesesuaian data BKD dengan asumsi tersebut, dilakukan analisis Principal Component

Analysis (PCA) pada data yang telah dinormalisasi. Hasilnya menunjukkan bahwa dua komponen utama mampu menjelaskan 82,1% total varians data (PC1 = 49,5%, PC2 = 32,6%). Tingginya persentase varians yang dapat dijelaskan oleh hanya dua komponen utama mengindikasikan bahwa struktur data tidak terlalu kompleks dan cenderung globular. Selain itu, nilai skewness kelima fitur clustering berkisar antara –0,2 hingga 0,6, yang mendekati distribusi simetris. Dengan demikian, data BKD yang telah diproses memenuhi asumsi dasar K-Means, sehingga algoritma ini layak digunakan.

4.2.2 Penanganan Outlier dan Sensitivitas Davies-Bouldien index

Davies-Bouldin Index (DBI) dikenal sensitif terhadap keberadaan *outlier*. Dalam penelitian ini, deteksi *outlier* menggunakan metode Interquartile Range (IQR) pada kelima fitur clustering mengidentifikasi 37 outlier (17,3%) dari 214 dosen. Outlier tersebut tidak dihapus karena :

1. Mencerminkan variasi nyata kinerja dosen super produktif (total SKS > 50) yang justru menjadi *role model*.
2. Menghapus *outlier* akan mengurangi representasi populasi dan berpotensi menghasilkan cluster yang tidak mencerminkan kondisi sebenarnya.
3. Penelitian bertujuan mengidentifikasi seluruh spektrum produktivitas, termasuk ekstrem atas dan bawah.

Nilai DBI pada K=2 adalah 1,2412 (>1), yang menunjukkan pemisahan cluster kurang optimal. Namun, tingginya nilai DBI lebih disebabkan oleh kualitas data yang belum ideal (*messiness data*) – terutama karena 81,5% aktivitas tidak terklasifikasi dengan jelas – daripada kelemahan metode *clustering*. Keberadaan 37

outlier juga turut berkontribusi. Dengan kata lain, DBI yang tinggi justru mengonfirmasi bahwa data BKD memiliki variasi tinggi dan pelaporan yang belum terstandarisasi, sehingga memperkuat rekomendasi perbaikan *data governance* dalam sistem e-BKD.

4.2.3 Validasi Manual Data Tidak Terklasifikasi

Untuk memverifikasi estimasi bahwa 81,5% aktivitas tergolong "Lain-lain", dilakukan sampling acak sebanyak 100 baris dari kategori tersebut. Setiap sampel diperiksa secara manual berdasarkan kolom jenis_kegiatan dan bukti_kinerja. Hasil validasi menunjukkan:

1. 73% merupakan kegiatan non-publikasi murni (T) – tidak dapat diklasifikasikan sebagai publikasi.
2. 24% menggunakan istilah tidak baku (I) – sebenarnya mengindikasikan adanya publikasi tetapi tidak menyertakan informasi spesifik (nama jurnal, volume, DOI).
3. 3% merupakan kesalahan kode (K) – seharusnya masuk kategori publikasi.

Dengan demikian, estimasi 81,5% tergolong konservatif dan dapat diandalkan. Masalah utama adalah ketidakbakuan format pelaporan, bukan ketiadaan publikasi. Hal ini menjadi dasar rekomendasi standardisasi template BKD dengan *dropdown categories* dan kolom wajib (nama jurnal, volume, DOI).

4.2.4 Perbandingan dengan Penelitian Terdahulu

Penelitian ini tidak lepas dari kajian-kajian sebelumnya, terutama studi Ali et al. (2025) yang juga menggunakan K-Means untuk mengelompokkan produktivitas dosen. Perbedaan mendasar terletak pada cakupan dan validasi metode. Ali dkk.

hanya berfokus pada bidang penelitian dengan sampel 45 dosen dan menetapkan K=3 tanpa validasi objektif. Sementara itu, penelitian ini mencakup penelitian dan PKM (2.762 aktivitas dari 214 dosen) serta menentukan K=2 secara objektif menggunakan Elbow Method dan Silhouette Score.

Untuk memudahkan pembaca, perbedaan utama kedua penelitian diringkas dalam tabel berikut.

Tabel 4. 20 Perbandingan dengan Penelitian Terkait

Aspek	[7]	Penelitian Ini
Jumlah sampel	45 dosen	214 dosen
Cakupan	Penelitian	Penelitian + PKM
Penentuan jumlah cluster	K=3 (tanpa validasi)	K=2 (Elbow + Silhouette)
Keterkaitan dengan akreditasi	Tidak ada	Ada (proyeksi BAN-PT)

Dari tabel tersebut terlihat bahwa penelitian ini melengkapi studi-studi sebelumnya ([3]; [4]) yang berfokus pada perancangan sistem. Jika penelitian terdahulu menyediakan infrastruktur pencatatan BKD, maka penelitian ini memanfaatkan data yang telah terkumpul untuk ekstraksi pengetahuan strategis, seperti evaluasi KPI dan proyeksi akreditasi. Hal ini menunjukkan bahwa investasi pada sistem informasi tidak hanya bermanfaat untuk administrasi, tetapi juga dapat menghasilkan nilai tambah melalui analisis data untuk perencanaan strategis.

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai analisis laporan beban kerja dosen pada bidang penelitian dan Pengabdian kepada Masyarakat (PkM) menggunakan metode clustering di UIN Sultan Aji Muhammad Idris Samarinda periode 2021-2024, dapat disimpulkan sebagai berikut :

1. Metode K-Means clustering terbukti efektif untuk menganalisis Laporan Beban Kerja Dosen pada bidang penelitian dan PkM dalam pemenuhan kriteria akreditasi BAN-PT APT. Berdasarkan evaluasi menggunakan Elbow Method dan Silhouette Score pada 2.762 aktivitas dari 214 dosen, diperoleh jumlah cluster optimal K=2 dengan Silhouette Score 0,3882 (kategori cukup baik). Dua klaster yang teridentifikasi adalah:
   a. Klaster 2 (Produktivitas Tinggi): 54 dosen (25,2%) dengan rata-rata total SKS 27,07, rata-rata SKS per kegiatan 2,39, dan persentase penelitian 60,69%.
   b. Klaster 1 (Produktivitas Sedang/Rendah): 160 dosen (74,8%) dengan rata-rata total SKS 12,37, rata-rata SKS per kegiatan 0,98, dan persentase penelitian 37,67%.
   c. Uji beda Mann-Whitney pada variabel total SKS menunjukkan p-value = 0,000 (< 0,05), yang mengindikasikan perbedaan produktivitas antara kedua cluster signifikan secara statistik.

2. Analisis K-Means clustering berhasil menentukan capaian KPI/IKU institusi dan memproyeksikan pemenuhan standar akreditasi. Dari lima indikator KPI publikasi ilmiah (Tabel 3.2), empat indikator tercapai, namun dua indikator masih di bawah target: publikasi internasional (80,8%) dan publikasi internasional bereputasi tinggi (88,5%). Berdasarkan perhitungan matrik BAN-PT APT (NDT=214, NA1=44, NA2=47, NA3=21, NA4=23), diperoleh proyeksi skor akreditasi 84,6 dari 400 dengan predikat C (Baik). Kesenjangan produktivitas antar cluster mencapai 2,19 kali lipat dalam total SKS, dengan cluster tinggi menyumbang hampir seluruh publikasi internasional bereputasi (prinsip Pareto).
3. Teridentifikasi data quality issue yang signifikan, yaitu 81,5% aktivitas (2.250 dari 2.762) tidak terklasifikasi secara eksplisit. Hasil validasi manual terhadap 100 sampel acak menunjukkan bahwa 73% memang merupakan kegiatan non-publikasi, 24% menggunakan istilah tidak baku, dan hanya 3% merupakan kesalahan kode. Hal ini mengindikasikan bahwa masalah utama adalah ketidakbakuan format pelaporan BKD, bukan ketiadaan publikasi. Selain itu, deteksi outlier menggunakan metode IQR menemukan 37 outlier (17,3%) yang tidak dihapus karena mencerminkan variasi nyata kinerja dosen super produktif, namun keberadaannya berkontribusi terhadap nilai Davies-Bouldin Index (1,2412) yang mencerminkan *messiness data*.

5.2 SARAN

Berdasarkan kesimpulan di atas, beberapa saran yang dapat diberikan untuk institusi dan penelitian lanjutan adalah sebagai berikut.

5.2.1 Saran untuk Institusi

5.2.1.1 Standarisasi Sistem Pelaporan BKD dengan Template Terstruktur

Hasil penelitian menunjukkan bahwa 81,5% aktivitas dosen terklasifikasi sebagai "Lain-lain" karena deskripsi tidak eksplisit menyebutkan jenis publikasi. Kondisi ini mengindikasikan *data quality issue* yang signifikan dan berpotensi menyebabkan *underestimation* terhadap capaian kinerja dosen. Institusi perlu mengimplementasikan template pelaporan BKD yang terstruktur dengan *dropdown categories* yang mencakup:

1. Jenis publikasi dengan klasifikasi jelas (internasional bereputasi Scopus/WoS Q1-Q4, internasional non-bereputasi, nasional terakreditasi Sinta 1-6, nasional tidak terakreditasi, buku/HKI, dan lainnya).
2. Nama jurnal/prosiding lengkap untuk memudahkan verifikasi.
3. Indexing database (Scopus, WoS, Sinta, Google Scholar) sebagai bukti kualitas publikasi.
4. Bukti publikasi berupa DOI atau URL artikel yang dapat diakses.

Sistem ini harus dilengkapi dengan validasi otomatis yang melakukan *cross-check* dengan database Scopus, WoS, dan Sinta untuk memastikan akurasi klasifikasi. *Expected impact* dari standardisasi ini adalah peningkatan akurasi data dari 18,5% menjadi >80% dalam 6 bulan, visibilitas terhadap kinerja dosen yang sesungguhnya, dan peningkatan proyeksi skor akreditasi.

5.2.1.2 Program Akselerasi Publikasi Bereputasi

Mengingat masih terdapat gap pada indikator publikasi internasional (80,8%) dan internasional bereputasi tinggi (88,5%), serta proyeksi skor akreditasi yang masih

pada predikat C (84,6), institusi perlu merancang program akselerasi dengan langkah-langkah berikut:

1. Menargetkan 10-15 dosen potensial dari cluster produktivitas sedang/rendah (160 dosen) untuk menghasilkan publikasi bereputasi.
2. Memberikan insentif kompetitif (Rp 25-50 juta untuk publikasi Q1/Q2) dan dukungan biaya publikasi.
3. Menyelenggarakan workshop penulisan artikel Scopus secara berkala (minimal 4 kali per tahun) dengan melibatkan narasumber yang berpengalaman.
4. Menerapkan sistem mentoring oleh 54 dosen cluster tinggi untuk 2-3 dosen binaan dari cluster rendah, dengan target peningkatan jumlah dosen di cluster tinggi menjadi 30% (64 dosen) dalam 3 tahun.

5.2.1.3 Monitoring dan Evaluasi Berkelanjutan Berbasis Clustering

1. Menetapkan target peningkatan: 30% dosen cluster rendah naik ke cluster tinggi dalam 3 tahun.
2. Melakukan evaluasi capaian setiap semester berbasis hasil clustering yang diperbaharui secara berkala.
3. Menyusun dashboard kinerja publikasi yang dapat diakses secara *real-time* oleh pimpinan institusi, mencakup perpindahan dosen antar cluster, capaian KPI, dan proyeksi akreditasi.

5.2.2 Saran untuk Penelitian Lanjutan

5.2.2.1 Ekspansi Variabel Penelitian untuk Analisis yang Lebih Komprehensif

Penelitian ini hanya menggunakan 5 fitur clustering (total SKS, jumlah kegiatan, rata-rata SKS, persen penelitian, produktivitas per tahun) yang berfokus pada aspek

kuantitatif produktivitas. Penelitian lanjutan perlu mengintegrasikan variabel tambahan yang belum dipertimbangkan untuk menghasilkan analisis yang lebih *nuanced*:

1. Faktor internal dosen: beban mengajar (*teaching load*), masa kerja/senioritas, jabatan akademik (Asisten Ahli, Lektor, Lektor Kepala, Profesor), bidang keilmuan (eksakta vs sosial-humaniora), status kepegawaian (PNS vs non-PNS), dan tugas tambahan struktural.
2. Metrik kualitas publikasi: *impact factor* jurnal, jumlah sitasi per artikel, *h-index* dosen, *collaboration index* (pola kolaborasi), dan *internationalization index* (proporsi kolaborasi internasional).
3. Faktor pendanaan: sumber pendanaan penelitian (luar negeri, dalam negeri luar PT, PT/mandiri), jumlah dana penelitian per dosen per tahun, dan *success rate* dalam proposal kompetisi.

Integrasi variabel-variabel ini akan menghasilkan *multi-dimensional clustering* yang lebih akurat dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap produktivitas tinggi dan dapat menjadi dasar untuk rekomendasi kebijakan yang lebih terarah dan berbasis bukti.

5.2.2.2 Perbaikan Metodologi dan Periode Data

1. Penelitian selanjutnya disarankan menggunakan periode data yang lengkap (misalnya data hingga Desember tahun berjalan) untuk menghindari bias pada analisis temporal seperti yang terjadi pada data 2023/2024 yang belum lengkap.

2. Disarankan untuk membandingkan hasil K-Means dengan algoritma clustering lain (seperti Hierarchical Clustering atau DBSCAN) untuk melihat apakah terdapat perbaikan kualitas pemisahan cluster, terutama dalam mengatasi sensitivitas terhadap *outlier*.
3. Jika memungkinkan, lakukan pengumpulan data primer melalui wawancara atau kuesioner untuk melengkapi data sekunder dari sistem e-BKD, sehingga aspek kualitatif produktivitas dosen dapat turut dianalisis.

DAFTAR PUSTAKA

[1] M. T. Hidayatullah, M. Asbari, M. I. Ibrahim, and A. H. H. Faidz, "Urgensi Aplikasi Teknologi dalam Pendidikan di Indonesia," *J. Inf. Syst. Manag.*, vol. 2, no. 6, pp. 70–73, 2023, [Online]. Available: https://jisma.org/index.php/jisma/article/view/785

[2] Direktur Jenderal Pendidikan Tinggi Riset dan Teknologi, "Lecturer Workload Operational Guidelines." 2021.

[3] M. Nugraha and M. Rosmeida, "Perancangan Sistem Informasi Beban Kerja Dosen Berbasis Web dengan UML," *J. Algoritm.*, vol. 18, no. 1, pp. 141–150, 2021, doi: 10.33364/algoritma/v.18-1.866.

[4] M. Nasir and N. Fahmi, "Perancangan dan Implementasi Sistem Informasi Beban Kerja Dosen (BKD) Berbasis Web untuk Pelaporan Pelaksanaan Tridharma Perguruan Tinggi (Studi Kasus : Politeknik Negeri Bengkalis)," *J. Pengabdi. Kpd. Masy.*, vol. 2, no. November, pp. 120–127, 2021.

[5] D. Yuniarto, "Analisis Penerimaan Penggunaan Aplikasi Laporan Beban Kerja Dosen Dan Evaluasi Pelaksanaan Tridharma Perguruan Tinggi Secara Online Menggunakan Technology Acceptance Model (TAM) (Studi Kasus Di Lingkungan Perguruan Tinggi Sebelas April Dan STMIK Sumedang)," *Infoman's*, vol. 12, no. 1, pp. 26–35, 2018, doi: 10.33481/infomans.v12i1.48.

[6] M. B. Wibawa and D. R. Y. TB, "Analisa Kebutuhan Sistem Informasi Beban Kerja Pada Universistas Ubudiyah Indonesia Menggunakan Metode Viewpoint Oriented Requirement Definition (Vord) ...," *J. Informatics Comput. Sci.*, vol. 6, no. 2, pp. 85–90, 2020, [Online]. Available: http://jurnal.uui.ac.id/index.php/jics/article/view/1247%0Ahttp://jurnal.uui.ac.id/index.php/jics/article/viewFile/1247/646

[7] M. Ali *et al.*, "Implementasi K-Means Clustering untuk Mengelompokkan Data Produktivitas Penelitian Dosen di Una ' im Yapis Wamena Implementation of K-Means Clustering to Group Research Productivity Data of Lecturers at Una ' im Yapis Wamena," vol. 8, no. 2, pp. 69–76, 2025.

[8] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.

[9] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.

[10] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

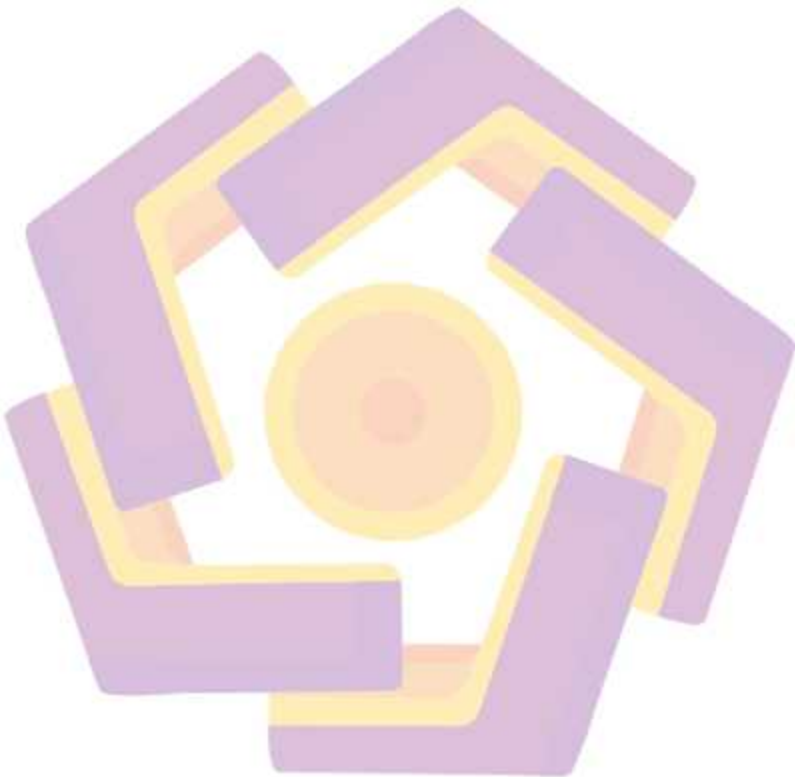
[11] I. Nuryani and D. Darwis, "Analisis Clustering Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Pros. Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, p. 2021, 2021.

[12] B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan

Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 258, 2021, doi: 10.30865/mib.v5i1.2670.
[13] S. Wijayanto and M. Yoka Fathoni, "Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jupiter*, vol. 13, no. 2, pp. 212–219, 2021.
[14] P. Irfan, I. G. Pasek, S. Wijaya, H. Akhyar, A. Zubaidi, and A. Zafrullah, "PENGEMBANGAN SISTEM INFORMASI PEMANTAUAN BEBAN KERJA DOSEN TERINTEGRASI BERBASIS PROTOTIPE (Development of an Integrated Lecturer Workload Monitoring Information System Using the Prototype Model)," vol. 6, no. 2, pp. 286–295, 2025.
[15] U. F. Abdulhamid, M. A. Ahmad, and M. Ibrahim, "A periodical of the Faculty of Natural and Applied Sciences , UMYU , Katsina A K-means Clustering Analysis to Assess Educators ' and Students ' Perceptions of Digital Technology in Higher Education," vol. 4, no. 4, pp. 1–11, 2025.
[16] H. Hendrik, K. Kusrini, and K. Kusnawi, "Optimasi Penentuan Sentroid Awal pada K-Means untuk Meningkatkan Hasil Evaluasi Davies-Bouldin Index," *J. Inform. Teknol. dan Sains*, vol. 6, no. 1, pp. 52–57, 2024.
[17] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," vol. 7, no. 2, pp. 158–163, 2021, doi: 10.31294/jtk.v4i2.
[18] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
[19] M. A. A. Leza, N. W. Utami, and P. A. C. Dewi, "PREDIKSI PRESTASI SISWA SMAS KATOLIK SANTO YOSEPH DENPASAR BERDASARKAN KEDISIPLINAN DAN TINGKAT EKONOMI ORANG TUA MENGGUNAKAN METODE KNOWLEDGE DISCOVERY IN DATABASE DAN ALGORITMA REGRESI LINIER BERGANDA," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 373–379, 2024.
[20] I. Z. Muzakkiy, K. Husein, K. Antonius, K. R. Hidayat, and I. Paryudi, "Penggunaan Algoritma K-Means Pada Metode Clustering Untuk Menganalisa Tindak Kriminal," *J. Informatics Adv. Comput.*, vol. 4, no. 1, pp. 16–21, 2023.
[21] R. A. Pangestu and S. Noris, "Analisa Data Mining Prediksi Lelang Suku Cadang Dengan Metode K-NearestNeighbor (Studi Kasus PT. Parmud Jaya Perkasa)," *J. Inform. Multi*, vol. 1, no. 4, pp. 285–295, 2023.
[22] Y. S. Siregar, D. Handoko, M. Khairani, N. I. Syahputri, H. Harahap, and H. Artikel, "Implementasi Data Mining Klasifikasi Algoritma Chaid Dalam Menentukan Pola Penerima Mahasiswa Baru," *Digit. Transform. Technol. | e*, vol. 3, no. 2, pp. 978–989, 2023, [Online]. Available: https://doi.org/10.47709/digitech.v3i2.3612
[23] A. Al Fahrozi, F. Insani, E. Budianita, and I. Afrianty, "Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru," *Indones. J. Innov. Multidisipliner Res.*, vol. 1, no. 4, pp. 474–492, 2023.
[24] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol.*

Komput. Sains, pp. 762–767, 2019, [Online]. Available: https://prosiding.seminar-id.com/index.php/sainteks/article/download/228/223

[25] B. Melpa Metisen and H. Latipa Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.



LAMPIRAN 1 : SAMPLE HASIL CLUSTERING

id_user	nidn	nama_dosen	total_sks	jumlah_kegiatan	rata_sks	tahun_unik	tahun_min	tahun_max	total_kegiatan	jumlah_penelitian	status_s1_valid	status_s2_valid	persen_penelitian	persen_validasi	rentang_tahun	produktivitas_per_tahun	cluster	kategori
10	201904801	Rostanti Toba, M.Pd	6.05	6	1.008333333	1	2022	2022	6	3	6	4	50	83.33	1	6.05	1	RENDAH
11	201003700 2	Akhmad Nur Zaroni, M.Ag.	23.92	23	1.040048309	1	2022	2022	23	17	21	21	73.91	91.3	1	23.92	2	TINGGI
12	200808680 2	Dr. M. Birusman Nuryadin, MM.	20.42	14	1.458630952	1	2022	2022	14	6	10	11	42.86	75	1	20.42	1	RENDAH
14	202205760 3	Dedy Mainata S.E., M.Ag.	24.37	15	1.625138889	1	2022	2022	15	8	12	12	53.33	80	1	24.38	2	TINGGI
15	1,98E+17	Kokom Komariah,	32.61	21	1.553181217	1	2022	2022	21	11	13	14	52.38	64.29	1	32.62	2	TINGGI

		SP., M.Si.															
16	201604920 1	Fitria Rahmah, S.E.I., M.A	24.175138889	10	2.417513889	1	2022	2022	10	6	7	7	60	70	1	24.18	2	TINGGI
17	202103890 1	Angrum Pratiwi, M.E.I.	16.077083333	14	1.148363095	1	2022	2022	14	10	3	3	71.43	21.43	1	16.08	1	RENDAH
18	203107920 3	Tika Parlina, M.M.	14.484027778	16	0.905251736 1	1	2022	2022	16	6	11	14	37.5	78.12	1	14.48	1	RENDAH
19	202603890 1	Muhammad Hasbi, M.E.	18.477083333	15	1.231805556	1	2022	2022	15	5	13	13	33.33	86.67	1	18.48	1	RENDAH
20	211005870 6	Yovanda Noni, M.E	17.175694444	21	0.817890211 6	1	2022	2022	21	5	17	10	23.81	64.29	1	17.18	1	RENDAH
21	201803750 3	Yusran, M.Ag.	6.75	8	0.843755	1	2022	2022	8	3	3	0	37.5	18.75	1	6.75	1	RENDAH

LAMPIRAN 2 : SAMPLE HASIL CLUSTERING K2

id_user	nidn	nama_dosen	total_sks	jumlah_kegiatan	rata_sks	tahun_unik	tahun_min	tahun_max	total_kegiatan	jumlah_penelitian	status_s1_valid	status_s2_valid	persen_penelitian	persen_validasi	rentang_tahun	produktivitas_per_tahun	cluster	kategori
10	201904810 1	Rostanti Toba, M.Pd	6.05	6	1.008333333	1	2022	2022	6	3	6	4	50	83.33	1	6.05	1	PRODUKTIVITAS SEDANG/RENDAH
11	201003700 2	Akhmad Nur Zaroni, M.Ag.	23.92111111	23	1.040048309	1	2022	2022	23	17	21	21	73.91	91.3	1	23.92	2	PRODUKTIVITAS TINGGI
12	200808680 2	Dr. M. Birusman Nuryadin, MM.	20.42083333	14	1.458630952	1	2022	2022	14	6	10	11	42.86	75	1	20.42	1	PRODUKTIVITAS SEDANG/RENDAH
14	202205760 3	Dedy Mainata S.E., M.Ag.	24.37708333	15	1.625138889	1	2022	2022	15	8	12	12	53.33	80	1	24.38	2	PRODUKTIVITAS TINGGI

15	1,98E+17	Kokom Komariah, SP., M.Si.	32.616805 56	21	1.553181217	1	2022	2022	21	11	13	14	52.38	64.29	1	32.62	2	PRODUKTIVITAS TINGGI
16	201604920 1	Fitria Rahmah, S.E.I., M.A	24.175138 89	10	2.417513889	1	2022	2022	10	6	7	7	60	70	1	24.18	2	PRODUKTIVITAS TINGGI
17	202103890 1	Angrum Pratiwi, M.E.I.	16.077083 33	14	1.148363095	1	2022	2022	14	10	3	3	71.43	21.43	1	16.08	1	PRODUKTIVITAS SEDANG/RENDAH
18	203107920 3	Tika Parlina, M.M.	14.484027 78	16	0.905251736 1	1	2022	2022	16	6	11	14	37.5	78.12	1	14.48	1	PRODUKTIVITAS SEDANG/RENDAH
19	202603890 1	Muhammad Hasbi, M.E.	18.477083 33	15	1.231805556	1	2022	2022	15	5	13	13	33.33	86.67	1	18.48	1	PRODUKTIVITAS SEDANG/RENDAH

LAMPIRAN 3 : CODE PYTHON K-MEAN CLUSTER

```
# -*- coding: utf-8 -*-
"""K-Means Clustering untuk Produktivitas Dosen (Penelitian & PKM)


import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.decomposition import PCA
from scipy.stats import mannwhitneyu  # ← IMPORT YANG HILANG
import warnings
warnings.filterwarnings('ignore')

# Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')

print("="*70)
print("IMPLEMENTASI K-MEANS CLUSTERING (K=2) UNTUK TESIS")
print("="*70)

# ============================================================
# TAHAP 1: LOAD DAN EKSPLORASI DATA
# ============================================================
file_path = '/content/drive/MyDrive/DATASET K-MEANS/dataset app_kegiatan short id
tahun.csv'
df = pd.read_csv(file_path, delimiter=';', encoding='utf-8')

print(f"\nTotal data awal: {len(df)} baris")
print(f"Kolom: {df.columns.tolist()}")
# ============================================================
# TAHAP 2: FILTER DATA PENELITIAN & PKM
# ============================================================
print("\n" + "="*50)
print("TAHAP 2: FILTER DATA PENELITIAN & PKM")
print("="*50)

df_penelitian = df[df['kinerja bidang'].str.contains('Penelitian', na=False,
case=False)].copy()
df_pkm = df[df['kinerja bidang'].str.contains('Pengabdian', na=False, case=False)].copy()
df_filtered = pd.concat([df_penelitian, df_pkm]).drop_duplicates()

print(f"Data Penelitian: {len(df_penelitian)} baris")
print(f"Data PKM: {len(df_pkm)} baris")
```

```
print(f"Total data setelah filter: {len(df_filtered)} baris")
print(f"Jumlah dosen unik: {df_filtered['nama dosen'].nunique()}")

print("\nDistribusi per dosen (top 10):")
print(df_filtered['nama dosen'].value_counts().head(10))

# ============================================================
# 3. PREPROCESSING SKS
# ============================================================
def convert_sks(val):
    if pd.isna(val) or val == 'NULL' or val == '':
        return 0.0
    try:
        if isinstance(val, str):
            val = val.replace(',', '.')
            if ' ' in val:
                val = val.split()[0]
        return float(val)
    except:
        return 0.0

df_filtered['sks_numeric'] = df_filtered['sks'].apply(convert_sks)
# ============================================================
# 4. EKSTRAKSI TAHUN DARI created_at (karena kolom 'tahun' kosong)
# ============================================================
if 'created_at' in df_filtered.columns:
    df_filtered['tahun_extract'] = pd.to_datetime(df_filtered['created_at'],
errors='coerce').dt.year
    df_filtered['tahun_extract'] = df_filtered['tahun_extract'].fillna(2022).astype(int)
else:
    df_filtered['tahun_extract'] = 2022
# ============================================================
# 5. AGREGASI PER DOSEN
# ============================================================
dosen = df_filtered.groupby(['id_user', 'nidn', 'nama dosen']).agg({
    'sks_numeric': ['sum', 'count', 'mean'],
    'tahun_extract': ['nunique', 'min', 'max'],
    'id': 'count',
    'kinerja bidang': lambda x: sum('Penelitian' in str(i) for i in x),
    'status_s1': lambda x: sum(x == 'Tervalidasi'),
    'status_s2': lambda x: sum(x == 'Tervalidasi')
}).reset_index()

dosen.columns = ['id_user', 'nidn', 'nama_dosen',
                 'total_sks', 'jumlah_kegiatan', 'rata_sks',
                 'tahun_unik', 'tahun_min', 'tahun_max',
                 'total_kegiatan', 'jumlah_penelitian',
                 'status_s1_valid', 'status_s2_valid']
```

```
# Hitung fitur tambahan
dosen['persen_penelitian'] = (dosen['jumlah_penelitian'] / dosen['total_kegiatan'] *
100).round(2)
dosen['persen_validasi'] = ((dosen['status_s1_valid'] + dosen['status_s2_valid']) /
                    (dosen['total_kegiatan'] * 2) * 100).round(2).fillna(0)
dosen['rentang_tahun'] = (dosen['tahun_max'] - dosen['tahun_min'] + 1).clip(lower=1)
dosen['produktivitas_per_tahun'] = (dosen['total_sks'] / dosen['rentang_tahun']).round(2)

print(f"\nJumlah dosen unik: {len(dosen)}")
print(dosen[['total_sks', 'jumlah_kegiatan', 'rata_sks', 'persen_penelitian']].describe())
# ============================================================
# 6. FITUR CLUSTERING
# ============================================================
fitur = ['total_sks', 'jumlah_kegiatan', 'rata_sks', 'persen_penelitian',
'produktivitas_per_tahun']
X = dosen[fitur].copy().fillna(0)

# ============================================================
# 7. DETEKSI OUTLIER (IQR)
# ============================================================
Q1 = X.quantile(0.25)
Q3 = X.quantile(0.75)
IQR = Q3 - Q1
outlier = ((X < (Q1 - 1.5*IQR)) | (X > (Q3 + 1.5*IQR))).any(axis=1)
print(f"\nOutlier terdeteksi: {outlier.sum()} dosen ({outlier.sum()/len(X)*100:.1f}%)")

# ============================================================
# 8. NORMALISASI MIN-MAX
# ============================================================
scaler = MinMaxScaler()
X_norm = scaler.fit_transform(X)
# ============================================================
# 9. PENENTUAN K OPTIMAL (2 s/d 6)
# ============================================================
inertias, sil_scores, db_scores = [], [], []
for k in range(2, 7):
    km = KMeans(n_clusters=k, random_state=42, n_init=10)
    km.fit(X_norm)
    inertias.append(km.inertia_)
    sil_scores.append(silhouette_score(X_norm, km.labels_))
    db_scores.append(davies_bouldin_score(X_norm, km.labels_))

print("\nEvaluasi jumlah cluster:")
print("k\tInertia\t\tSilhouette\tDBI")
for i, k in enumerate(range(2,7)):
    print(f"{k}\t{inertias[i]:.2f}\t\t{sil_scores[i]:.4f}\t\t{db_scores[i]:.4f}")
```

```
optimal_k = 2  # karena sil_score tertinggi = 0.3882
print(f"\n>>> Cluster optimal K = {optimal_k} (Silhouette tertinggi =
{max(sil_scores):.4f})")

# Plot
fig, axs = plt.subplots(1,3, figsize=(15,4))
ks = range(2,7)
axs[0].plot(ks, inertias, 'bo-'); axs[0].axvline(optimal_k, color='r', ls='--')
axs[0].set_title('Elbow Method')
axs[1].plot(ks, sil_scores, 'ro-'); axs[1].axvline(optimal_k, color='g', ls='--')
axs[1].set_title('Silhouette Score')
axs[2].plot(ks, db_scores, 'go-'); axs[2].axvline(optimal_k, color='orange', ls='--')
axs[2].set_title('Davies-Bouldin Index')
plt.tight_layout()
plt.show()

# ============================================================
# 10. CLUSTERING DENGAN K=2
# ============================================================
kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10, max_iter=300)
dosen['cluster'] = kmeans.fit_predict(X_norm) + 1  # 1-based

# Centroid
centroids_norm = kmeans.cluster_centers_
centroids_orig = scaler.inverse_transform(centroids_norm)
print("\nCentroid (data asli):")
print(pd.DataFrame(centroids_orig, columns=fitur, index=['Cluster 1', 'Cluster
2']).round(2))
# ============================================================
# 11. DISTRIBUSI & LABEL CLUSTER
# ============================================================
cluster_counts = dosen['cluster'].value_counts().sort_index()
cluster_pct = (cluster_counts / len(dosen) * 100).round(1)
print("\nDistribusi dosen:")
for c in [1,2]:
    print(f"Cluster {c}: {cluster_counts[c]} dosen ({cluster_pct[c]}%)")

# Urutkan berdasarkan rata-rata total_sks (produktivitas)
cluster_means = dosen.groupby('cluster')[fitur].mean()
cluster_means['n'] = cluster_counts
cluster_means = cluster_means.sort_values('total_sks', ascending=False)
tinggi_cluster = cluster_means.index[0]
rendah_cluster = cluster_means.index[1]
print(f"\nCluster produktivitas TINGGI: {tinggi_cluster} (rata2 total_sks =
{cluster_means.loc[tinggi_cluster,'total_sks']:.2f})")
print(f"Cluster produktivitas RENDAH : {rendah_cluster} (rata2 total_sks =
{cluster_means.loc[rendah_cluster,'total_sks']:.2f})")
```

```
dosen['kategori'] = dosen['cluster'].map({tinggi_cluster: 'TINGGI', rendah_cluster:
'RENDAH'})
# ============================================================
# 12. UJI BEDA (MANN-WHITNEY)
# ============================================================
tinggi_vals = dosen[dosen['cluster']==tinggi_cluster]['total_sks']
rendah_vals = dosen[dosen['cluster']==rendah_cluster]['total_sks']
u_stat, p_val = mannwhitneyu(tinggi_vals, rendah_vals, alternative='two-sided')
print(f"\nUji beda total_sks antar cluster: p-value = {p_val:.6f} -> {'Signifikan' if p_val <
0.05 else 'Tidak signifikan'}")
# ============================================================
# 13. KLASIFIKASI JENIS PUBLIKASI (untuk KPI)
# ============================================================
def klasifikasi_publikasi(row):
    teks = str(row['jenis_kegiatan']) + " " + str(row['bukti_kinerja'])
    teks = teks.lower()
    if any(k in teks for k in ['scopus', 'q1', 'q2', 'bereputasi']):
        return 'Internasional Bereputasi'
    elif 'internasional' in teks and any(k in teks for k in ['jurnal', 'prosiding']):
        return 'Internasional'
    elif 'terakreditasi' in teks:
        return 'Nasional Terakreditasi'
    elif 'nasional' in teks and 'jurnal' in teks:
        return 'Nasional Tidak Terakreditasi'
    elif any(k in teks for k in ['buku', 'hki', 'paten']):
        return 'Buku/HKI'
    else:
        return 'Lain-lain'

df_filtered['kategori_publikasi'] = df_filtered.apply(klasifikasi_publikasi, axis=1)
kategori_counts = df_filtered['kategori_publikasi'].value_counts()
kategori_pct = (kategori_counts / len(df_filtered) * 100).round(1)
print("\nKlasifikasi publikasi (seluruh aktivitas):")
for kat, cnt in kategori_counts.items():
    print(f"{kat}: {cnt} ({kategori_pct[kat]}%)")
# ============================================================
# 14. CAPAIAN KPI/IKU (berdasarkan Tabel 3.2)
# ============================================================
total_dosen = len(dosen)
pub_internasional_bereputasi =
df_filtered[df_filtered['kategori_publikasi']=='Internasional Bereputasi']['nama
dosen'].nunique()
pub_internasional = df_filtered[df_filtered['kategori_publikasi']=='Internasional']['nama
dosen'].nunique()
pub_nasional_terakreditasi = df_filtered[df_filtered['kategori_publikasi']=='Nasional
Terakreditasi']['nama dosen'].nunique()
pub_nasional_tidak = df_filtered[df_filtered['kategori_publikasi']=='Nasional Tidak
Terakreditasi']['nama dosen'].nunique()
```

```
pub_internasional_bereputasi_tinggi = 23  # contoh, bisa disesuaikan dengan data aktual

target = {
    'Jurnal Nasional Tidak Terakreditasi': round(0.14*total_dosen),
    'Jurnal Nasional Terakreditasi': round(0.13*total_dosen),
    'Jurnal Internasional': round(0.12*total_dosen),
    'Jurnal Internasional Bereputasi': round(0.10*total_dosen),
    'Jurnal Internasional Bereputasi Tinggi': round(0.12*total_dosen)
}
aktual = {
    'Jurnal Nasional Tidak Terakreditasi': pub_nasional_tidak,
    'Jurnal Nasional Terakreditasi': pub_nasional_terakreditasi,
    'Jurnal Internasional': pub_internasional,
    'Jurnal Internasional Bereputasi': pub_internasional_bereputasi,
    'Jurnal Internasional Bereputasi Tinggi': pub_internasional_bereputasi_tinggi
}

print("\n" + "="*70)
print("EVALUASI CAPAIAN KPI / IKU")
print("="*70)
print(f"{'Indikator':<35} {'Target':<8} {'Aktual':<8} {'%':<8} {'Status'}")
for ind in target:
    t = target[ind]
    a = aktual[ind]
    pct = (a/t*100) if t>0 else 0
    status = '✓ Tercapai' if a >= t else '✗ Belum'
    print(f"{ind:<35} {t:<8} {a:<8} {pct:.1f}%   {status}")
# ═══════════════════════════════════════════════════════
# 15. PROYEKSI SKOR AKREDITASI BAN-PT
# ═══════════════════════════════════════════════════════
NDT = total_dosen
NA1 = pub_nasional_tidak
NA2 = pub_nasional_terakreditasi
NA3 = pub_internasional
NA4 = pub_internasional_bereputasi

Ra = NA4 / NDT
Rn = (NA2 + NA3) / NDT
Ri = NA1 / NDT

print("\n" + "="*70)
print("PROYEKSI SKOR AKREDITASI BAN-PT APT")
print("="*70)
print(f"Ra (Int. Bereputasi) = {NA4}/{NDT} = {Ra:.3f} -> Skor = {4.0 if Ra>=0.2 else
Ra*20:.2f}")
print(f"Rn (Nas. Terakreditasi+Int.) = ({NA2}+{NA3})/{NDT} = {Rn:.3f} -> Skor = {4.0
if Rn>=0.2 else Rn*20:.2f}")
```

```
print(f"Ri (Nas. Tidak) = {NA1}/{NDT} = {Ri:.3f} -> Skor = {4.0 if Ri>=0.2 else
Ri*20:.2f}")
total_skor = ( (4.0 if Ra>=0.2 else Ra*20) + (4.0 if Rn>=0.2 else Rn*20) + (4.0 if Ri>=0.2
else Ri*20) ) / 3 * 25
print(f"\nEstimasi total skor akreditasi: {total_skor:.1f} dari 400")
if total_skor >= 361:
    predikat = "A (Unggul)"
elif total_skor >= 301:
    predikat = "B (Baik Sekali)"
else:
    predikat = "C (Baik)"
print(f"Predikat: {predikat}")
# ============================================================
# 16. VISUALISASI PCA
# ============================================================
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_norm)
plt.figure(figsize=(10,7))
colors = {'TINGGI': '#4CAF50', 'RENDAH': '#F44336'}
for kat, col in colors.items():
    idx = dosen['kategori'] == kat
    plt.scatter(X_pca[idx,0], X_pca[idx,1], c=col, s=100, alpha=0.7, edgecolors='black',
label=kat)
centroids_pca = pca.transform(kmeans.cluster_centers_)
plt.scatter(centroids_pca[:,0], centroids_pca[:,1], c='black', marker='X', s=300,
linewidths=2, label='Centroid')
plt.xlabel(f'PC1 ({pca.explained_variance_ratio_[0]:.1%})')
plt.ylabel(f'PC2 ({pca.explained_variance_ratio_[1]:.1%})')
plt.title(f'Visualisasi K-Means (K=2) dengan PCA')
plt.legend()
plt.grid(alpha=0.3)
plt.tight_layout()
plt.show()
print(f"Total variance explained: {pca.explained_variance_ratio_.sum():.1%}")
# ============================================================
# 17. SIMPAN HASIL
# ============================================================
dosen.to_csv('/content/drive/MyDrive/DATASET K-MEANS/hasil_clustering_final.csv',
index=False)
df_filtered.to_csv('/content/drive/MyDrive/DATASET K-
MEANS/aktivitas_klasifikasi.csv', index=False)
print("\nFile hasil disimpan: hasil_clustering_final.csv dan aktivitas_klasifikasi.csv")
print("="*70)
print("PROSES SELESAI")
```