

## PREDIKSI PELUANG LOLOS PROGRAM STUDI SIMAK UI TAHUN 2009 MENGGUNAKAN KLASTER K-MEANS DATA HISTORIS 2006-2008

Ahmad Fachriza<sup>1</sup>, Wahyu Tri Jayaputra<sup>2</sup>, Diaz Permana Putra<sup>3</sup>, Darris Daffa  
Muhammad<sup>4</sup>, Arjuna Rizki Kurnia<sup>5</sup>, Sartika Lina Mulani Sitio<sup>6</sup>

[ahmdfchrzz@gmail.com](mailto:ahmdfchrzz@gmail.com)<sup>1</sup>, [wahyutrijp@gmail.com](mailto:wahyutrijp@gmail.com)<sup>2</sup>, [diazpermanaputra123@gmail.com](mailto:diazpermanaputra123@gmail.com)<sup>3</sup>,  
[darisdafal@gmail.com](mailto:darisdafal@gmail.com)<sup>4</sup>, [arjunarizkikurnia1717@gmail.com](mailto:arjunarizkikurnia1717@gmail.com)<sup>5</sup>, [dosen00847@unpam.ac.id](mailto:dosen00847@unpam.ac.id)<sup>6</sup>

Universitas Pamulang

### ABSTRAK

Persaingan masuk perguruan tinggi melalui seleksi Simak UI menunjukkan peningkatan rasio peminat per daya tampung pada periode 2006–2008, yang memerlukan prediksi kuantitatif untuk strategi pendaftaran calon mahasiswa dan perencanaan kapasitas universitas. Penelitian ini bertujuan mengelompokkan 20 program studi Simak UI berdasarkan tingkat persaingan historis serta meramalkan persaingan tahun 2009 dengan algoritma K-Means ( $k = 3$ ). Data bersumber dari Sarwono (2010), meliputi jumlah peminat tahun 2006–2007, daya tampung tahun 2008, dan persentase persaingan tahunan. Tahapan Knowledge Discovery in Databases (KDD) diterapkan, mencakup imputasi nilai hilang, Min–Max Scaling, penentuan  $k$  optimal melalui elbow method dan silhouette score ( $\geq 0,5$ ), serta implementasi K-Means di Python (scikit-learn). Hasil clustering menghasilkan tiga kelompok karakteristik persaingan rendah (rata-rata 6,1 %), sedang (14,2 %), dan tinggi (24,8 %), dengan silhouette score rata-rata 0,62. Model regresi linier berbasis centroid memprediksi persaingan 2009 masing-masing sebesar 5,3 %, 14,4 %, dan 25,7 % ( $R^2 = 0,89$ ). Kontribusi penelitian ini adalah penyediaan peta kluster persaingan historis dan prediksi yang dapat digunakan calon mahasiswa untuk menyusun pilihan studi serta membantu universitas dalam kebijakan daya tampung yang lebih adaptif.

**Kata Kunci:** K-Means, Simak UI, Clustering, Prediksi Persaingan, Data Mining.

### ABSTRACT

*Competition to enter higher education through SimakUI selection shows an increase in the ratio of applicants per capacity in the period 2006–2008, which requires quantitative predictions for prospective student registration strategies and university capacity planning. This study aims to group 20 SimakUI study programs based on historical competition levels and predict competition in 2009 using the K-Means algorithm ( $k=3$ ). Data sourced from Sarwono (2010), including the number of applicants in 2006–2007, capacity in 2008, and annual competition percentage. The Knowledge Discovery in Databases (KDD) stages are applied, including missing value imputation, Min–Max Scaling, determining optimal  $k$  using the elbow method and silhouette score ( $\geq 0.5$ ), and implementing K-Means in Python (scikit-learn). The clustering results produced three groups of low competition characteristics (average 6.1%), medium (14.2%), and high (24.8%), with an average silhouette score of 0.62. The centroid-based linear regression model predicted 2009 competition of 5.3%, 14.4%, and 25.7%, respectively ( $R^2=0.89$ ). The contribution of this study is the provision of historical and predicted competition cluster maps that can be used by prospective students to organize study choices and assist universities in more adaptive capacity policies.*

**Keywords:** K-Means, Simak UI, Clustering, Prediksi Persaingan, Data Mining.

### PENDAHULUAN

Sejak diberlakukannya Simak UI sebagai jalur seleksi masuk perguruan tinggi negeri,

rasio peminat per daya tampung menunjukkan tren kenaikan dari tahun ke tahun, misalnya persentase peminat untuk Program Studi Teknik Informatika meningkat lebih dari 25 % antara 2006 dan 2008 (Sarwono, 2010). Kondisi tersebut menuntut calon mahasiswa untuk melakukan strategi pendaftaran yang lebih matang, sekaligus mendorong universitas merumuskan kebijakan daya tampung yang responsif terhadap lonjakan permintaan.

Berbagai penelitian telah memanfaatkan algoritma K-Means dalam konteks clustering dan forecasting seleksi masuk. Yunita (2018) menerapkan kerangka KDD untuk clustering data penerimaan mahasiswa baru, termasuk preprocessing dan evaluasi cluster menggunakan silhouette score untuk memastikan kualitas kelompok (Yunita, 2018, hlm. 112). Firdaus et al. (2025) memperkaya metodologi tersebut dengan pendekatan parallel computing untuk mempercepat waktu eksekusi pada dataset besar, namun tetap menjaga koherensi kluster. Selain itu, Rahmawati dan Arifin (2020) mengelompokkan siswa SMAN 8 Bandung berdasarkan keberhasilan SNMPTN, menegaskan efektivitas K-Means dalam memetakan preferensi masuk perguruan tinggi.

Berdasarkan latar belakang tersebut, penelitian ini dirancang untuk menjawab dua pertanyaan utama: (1) Bagaimana pengelompokan program studi Simak UI 2006– 2008 ke dalam kluster persaingan rendah, sedang, dan tinggi menggunakan algoritma K- Means?; (2) Sejauh mana model clustering tersebut dapat memprediksi rasio peminat per daya tampung Simak UI pada tahun 2009? Dengan demikian, tujuan penelitian adalah: (a) mengidentifikasi kluster program studi berdasarkan tingkat persaingan; dan

(b) meramalkan peluang lolos calon mahasiswa pada Simak UI 2009, sehingga memberikan wawasan strategis bagi pendaftar dan pembuat kebijakan.

## **METODE**

Metode penelitian ini mengikuti kerangka kerja Knowledge Discovery in Databases (KDD) untuk menganalisis data historis Simak UI tahun 2006 hingga 2008. Data yang digunakan bersumber dari buku Psikologi Remaja edisi revisi oleh Sarwono (2010), yang mencakup 20 program studi dengan variabel seperti daya tampung tahun 2008, jumlah peminat tahun 2006 dan 2007, serta persentase persaingan pada tahun-tahun tersebut. Tahap awal melibatkan deskripsi dataset yang memperjelas struktur dan isi data sebelum dilakukan pemrosesan lebih lanjut. Variabel-variabel numerik seperti peminat dan persaingan menjadi dasar utama dalam analisis ini untuk membentuk kluster program studi berdasarkan pola kompetisinya.

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum dimasukkan ke dalam algoritma K-Means. Langkah awal adalah penanganan missing value, khususnya pada kolom “Peminat 2006” untuk program studi Sistem Informasi, yang diisi menggunakan nilai median agar tidak mengganggu distribusi data. Setelah itu, dilakukan normalisasi dengan metode Min–Max Scaling agar semua variabel berada dalam rentang [0,1], menghindari dominasi fitur berskala besar. Reduksi dimensi sebenarnya tidak diperlukan karena jumlah fitur tergolong sedikit, namun Principal Component Analysis (PCA) dapat digunakan bila visualisasi dua dimensi dibutuhkan untuk membantu interpretasi.

Penentuan jumlah kluster optimal dilakukan melalui dua pendekatan, yakni Elbow Method dan Silhouette Analysis. Elbow Method digunakan untuk mengamati titik di mana penurunan nilai Within-Cluster Sum of Squares (WCSS) mulai tidak signifikan, sementara Silhouette Score digunakan untuk mengevaluasi kualitas pemisahan antar kluster. Algoritma K-Means diimplementasikan menggunakan Python 3.8 di Visual Studio Code, dengan

bantuan pustaka scikit-learn, pandas, serta matplotlib dan seaborn untuk visualisasi. Parameter penting seperti `init='k-means++'` dan `n_init=10` diatur untuk memastikan stabilitas hasil. Evaluasi hasil dilakukan melalui cluster profiling, perhitungan rata-rata tiap klaster, dan visualisasi dua dimensi yang memperlihatkan distribusi klaster berdasarkan fitur-fitur utama. Semua hasil dikemas secara sistematis dan divisualisasikan dalam bentuk grafik untuk mendukung interpretasi akhir.

## HASIL DAN PEMBAHASAN

Penelitian ini mengkaji penerapan metode K-Means clustering terhadap data pendaftaran Simak UI tahun 2006 hingga 2008 untuk memetakan pola kompetisi antarprogram studi. Tujuan utama dari penelitian ini adalah mengelompokkan program studi berdasarkan jumlah peminat, daya tampung, dan tingkat persaingan, serta memproyeksikan prediksi kompetisi pada tahun 2009. Dalam proses ini, visualisasi data menjadi alat penting untuk memahami distribusi dan pola yang muncul di tiap tahun. Data yang dianalisis melalui visualisasi scatter dan pairplot memberikan gambaran rinci tentang karakteristik masing-masing klaster dan perubahan dinamika pendaftaran dari waktu ke waktu.

Pada tahun 2007, visualisasi hubungan antara jumlah peminat dan persentase persaingan menunjukkan pembagian klaster yang cukup tegas. Cluster 0 terdiri dari program studi yang memiliki tingkat persaingan rendah, umumnya di bawah 10%, dan jumlah peminat yang juga relatif sedikit, yaitu di bawah 600 orang. Cluster 1 mencakup program-program dengan persaingan sedang, yaitu antara 10% hingga 20%, dan jumlah peminat antara 600 hingga 1.200 orang. Sementara itu, Cluster 2 mencerminkan program-program studi yang sangat kompetitif, dengan persaingan di atas 20% dan peminat lebih dari 1.200 orang. Distribusi ini menegaskan bahwa program studi di UI memiliki tingkatan kompetisi yang berbeda dan dapat dikelompokkan secara sistematis menggunakan metode clustering.

Tren dari tahun ke tahun memperlihatkan bahwa terjadi peningkatan yang konsisten dalam hal jumlah peminat dan tingkat persaingan. Pergeseran titik-titik pada scatter plot ke arah kanan atas menggambarkan bahwa banyak program studi mengalami lonjakan peminat. Hal ini menunjukkan adanya peningkatan minat masyarakat terhadap perguruan tinggi, khususnya Universitas Indonesia, yang memicu peningkatan persaingan dalam proses seleksi masuk. Selain itu, pola ini menjadi indikasi penting bagi pengelola universitas untuk menyesuaikan kebijakan daya tampung serta strategi penyelenggaraan ujian masuk agar tetap mampu menjangkau mahasiswa berkualitas.

Analisis lebih lanjut dilakukan dengan membandingkan jumlah peminat dan daya tampung pada tahun 2008. Visualisasi scatter yang menggunakan garis  $y = x$  sebagai referensi, membantu menyoroti apakah suatu program studi memiliki kapasitas daya tampung yang cukup atau mengalami kekurangan kursi. Program-program yang tergolong dalam Cluster 2 terlihat secara konsisten berada di atas garis referensi, yang menunjukkan bahwa jumlah peminat jauh melebihi daya tampung yang tersedia. Ini menandakan adanya tekanan kapasitas yang tinggi pada jurusan-jurusan favorit. Sebaliknya, Cluster 0 lebih banyak berada di bawah garis dan dekat dengan sumbu, mengindikasikan bahwa daya tampung relatif memadai atau bahkan melebihi permintaan.

Untuk memperkaya pemahaman akan struktur klaster, penelitian juga menyajikan pairplot seluruh variabel penting, termasuk peminat tahun 2006 dan 2007, daya tampung tahun 2008, dan persentase persaingan tahun 2007. Hasil dari visualisasi ini menunjukkan adanya pola pemisahan yang cukup kuat antara klaster, yang mengindikasikan validitas

pemilihan jumlah klaster sebanyak tiga kelompok. Cluster 1 memiliki distribusi peminat yang lebih tersebar, sedangkan Cluster 2 menunjukkan dominasi dalam variabel persaingan. Ini memberikan justifikasi bahwa pemetaan program studi berdasarkan metode K-Means cukup efektif dalam menggambarkan kondisi nyata yang terjadi di lapangan.

Prediksi persentase persaingan pada tahun 2009 dilakukan dengan pendekatan linier menggunakan rata-rata centroid dari masing-masing klaster dan tren pertumbuhan dari tahun-tahun sebelumnya. Dengan menggunakan regresi sederhana, hasil analisis menunjukkan bahwa Cluster 0 memiliki tingkat persaingan paling stabil, dengan rata-rata hanya meningkat dari 5,2% menjadi 5,3%. Cluster 1 memperlihatkan kenaikan dari 13,8% menjadi 14,4%, sedangkan Cluster 2 mengalami lonjakan signifikan dari 24,5% menjadi 25,7%. Hasil ini menunjukkan bahwa program-program dalam Cluster 2 semakin ketat dan menuntut persaingan yang jauh lebih tinggi dari tahun-tahun sebelumnya.

Visualisasi prediksi pada Gambar 6 memperjelas bahwa kecenderungan peningkatan persaingan paling tajam terjadi di Cluster 2, yang mencakup program-program studi favorit. Ini menjadi sinyal bagi calon pendaftar untuk lebih berhati-hati dan matang dalam menentukan pilihan program studi. Program-program yang tergolong ke dalam Cluster 2 diprediksi akan menjadi lebih selektif dan hanya bisa diakses oleh peserta dengan nilai ujian sangat tinggi. Sebaliknya, Cluster 0 memberikan opsi yang relatif aman bagi pendaftar yang ingin meningkatkan peluang diterima di universitas tanpa harus menghadapi kompetisi ekstrem.

Interpretasi lebih lanjut terhadap klaster mengungkapkan karakter unik masing-masing kelompok. Cluster 0 berisi program-program dengan rasio peminat terhadap daya tampung yang rendah. Contohnya adalah Ilmu Komputer dan Farmasi, yang meskipun cukup diminati, memiliki kapasitas kursi yang sebanding dengan permintaan. Ini menyebabkan peluang kelulusan peserta ujian di klaster ini cenderung tinggi. Program-program dalam kelompok ini cocok untuk peserta dengan persiapan standar yang menginginkan kestabilan dalam peluang kelulusan.

Cluster 1 dihuni oleh jurusan-jurusan dengan tingkat persaingan sedang. Biologi dan Kimia adalah dua contoh yang menggambarkan posisi klaster ini. Persentase persaingan berkisar sekitar 14%, yang artinya seleksi cukup ketat, namun masih dalam batas wajar. Kenaikan tingkat persaingan dari tahun ke tahun juga cukup moderat, memberikan kesempatan bagi peserta ujian untuk menyusun strategi pendaftaran berdasarkan kekuatan akademik yang dimiliki. Jurusan dalam kelompok ini masih cukup fleksibel untuk dijadikan pilihan utama maupun cadangan.

Sementara itu, Cluster 2 dihuni oleh program-program yang tergolong paling kompetitif. Pendidikan Dokter dan Teknik Komputer menjadi contoh nyata dari kelompok ini, yang pada tahun 2008 telah mencapai tingkat persaingan di atas 20% dan diprediksi akan terus meningkat. Program-program dalam klaster ini menarik banyak peminat, tetapi hanya memiliki daya tampung yang terbatas. Oleh karena itu, calon mahasiswa harus memiliki nilai yang tinggi dan strategi pendaftaran yang matang, termasuk mempersiapkan pilihan cadangan di klaster lain.

Ketiga klaster ini tidak hanya membantu memetakan risiko dan peluang dari sisi peserta ujian, tetapi juga memberikan wawasan penting bagi pengambil kebijakan universitas. Informasi ini dapat digunakan untuk menyesuaikan daya tampung, merancang sistem seleksi yang adil, serta menyusun kebijakan penerimaan mahasiswa berdasarkan pola minat dan kapasitas yang tersedia. Dengan pemahaman yang lebih dalam tentang distribusi

kompetisi ini, pihak universitas dapat membuat perencanaan akademik yang lebih strategis.

Selain itu, prediksi kompetisi untuk tahun 2009 dapat menjadi pedoman bagi siswa kelas 12 yang akan mendaftar ke perguruan tinggi. Dengan mengetahui tingkat persaingan berdasarkan klaster, calon peserta ujian bisa menilai kemungkinan diterima di jurusan yang diincar. Misalnya, siswa yang memiliki nilai tinggi dan persiapan matang bisa memilih program studi di Cluster 2, sementara yang memiliki nilai menengah bisa fokus pada jurusan di Cluster 1. Jurusan di Cluster 0 tetap menjadi pilihan yang logis bagi siswa yang ingin bermain aman.

Implikasi dari analisis ini juga sangat berguna bagi guru BK, lembaga bimbingan belajar, dan orang tua siswa dalam memberikan saran akademik kepada anak didik. Dengan data ini, saran yang diberikan tidak lagi bersifat spekulatif, melainkan berdasarkan analisis kuantitatif yang dapat diandalkan. Hal ini memungkinkan siswa membuat keputusan pendidikan yang lebih realistis dan terukur.

Terakhir, penerapan K-Means dalam konteks pemetaan persaingan masuk perguruan tinggi seperti Simak UI membuktikan bahwa metode ini efektif digunakan dalam bidang pendidikan. Selain mudah diimplementasikan dengan perangkat lunak seperti Python, hasil klasterisasi juga mampu menggambarkan kenyataan di lapangan secara akurat. Dengan demikian, pendekatan ini layak dipertimbangkan untuk diterapkan secara lebih luas dalam sistem seleksi masuk perguruan tinggi di Indonesia.

## **KESIMPULAN**

Penerapan metode K-Means pada data Simak UI 2006–2008 berhasil membentuk tiga cluster persaingan yang bermakna, dengan silhouette score rata-rata 0,62, menegaskan kualitas pemisahan antar kelompok program studi. Cluster pertama, yang mencakup program studi dengan persaingan rata-rata 6,1 % dan peminat sekitar 482 orang, menunjukkan titik masuk relatif rendah, sedangkan cluster kedua dan ketiga masing-masing memiliki rata-rata persaingan 14,2 % (1.034 peminat) dan 24,8 % (1.856 peminat), yang mencerminkan intensitas permintaan tinggi. Tren peningkatan jumlah peminat dan persaingan setiap tahunnya, dengan Compound Annual Growth Rate (CAGR) persaingan sebesar 5,7 % dari 2006 ke 2008 telah mengindikasikan dinamika kompetitif yang terus menguat dalam seleksi masuk perguruan tinggi (Sarwono, 2010).

Selanjutnya, model regresi linier berbasis centroid cluster memprediksi persaingan Simak UI 2009 dengan  $R^2$  0,89, yaitu sebesar 5,3 %, 14,4 %, dan 25,7 % untuk cluster rendah, sedang, dan tinggi, sehingga memberikan keandalan prediktif yang tinggi bagi pengguna data. Temuan ini tidak hanya memetakan pola persaingan historis secara kuantitatif tetapi juga menunjang perencanaan strategi pendaftaran dan kebijakan daya tampung yang lebih adaptif.

## **DAFTAR PUSTAKA**

- Dahnial, A. (2023). Implementation of K-Means clustering method to lecturers based on publications of national journals and accredited SINTA. *JEECS (Journal of Electrical Engineering & Computer Science)*, 8(1), 27–40. <https://doi.org/10.54732/jeeecs.v8i1.4>
- Firdaus, A. F., Hadi, T., & Desnelita, Y. (2025). Parallel computing pada clustering K-Means untuk analisis keketatan program studi SNBT 2023. *Komputika*, 14(1), 17–24. <https://doi.org/10.34010/hpm5tg52>
- Izzati, W. R., Rahayu, S. S., & Sulistiani, W. (2019). Analisis potensi asal sekolah pada jalur

- penerimaan mahasiswa baru di Universitas Lampung menggunakan algoritma K-Means. *Electrician: Jurnal Rekayasa & Teknologi Elektro*, 13(1), 7–12. <https://doi.org/10.23960/elc.v13n1.2087>
- Kurniawan, H., Defit, S., & Sumijan. (2020). Data mining menggunakan metode K-Means clustering untuk menentukan besaran uang kuliah tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80–89. <https://doi.org/10.52158/jacost.v1i2.102>
- Rahmawati, D., & Arifin, S. (2020). Penerapan algoritma K-Means untuk pengelompokan siswa lolos SNMPTN di SMAN 8 Bandung. *Jurnal Responsif: Riset Sains dan Informatika*, 2(2), 572–577. <https://doi.org/10.29207/resti.v2i2.308>
- Sarwono, S. W. (2010). *Psikologi remaja* (Edisi revisi). Rajawali Persada.
- Yunita, F. (2018). Penerapan data mining menggunakan algoritma K-Means clustering pada penerimaan mahasiswa baru (Studi kasus: Universitas Islam Indragiri). *Jurnal SISTEMASI*, 4(2), 110–118.