

## ANALISIS DATA JUDI ONLINE DI 5 PROVINSI INDONESIA DENGAN METODE K-MEANS DAN DECISION TREE

Muhammad Bayu Saputra✉, Nor Anisa

Fakultas Sains & Teknologi, Universitas Sari Mulia, Banjarmasin, Indonesia

Email: [muhammad.bayu@student.unism.ac.id](mailto:muhammad.bayu@student.unism.ac.id)

### ABSTRACT

*This study aims to analyze online gambling data detected in five provinces in Indonesia using the K-Means and Decision Tree methods. The data includes player counts, transaction values, and geographical distribution in West Java, Jakarta, Central Java, Banten, and East Java. The K-Means method was applied to cluster provinces based on player counts and transaction values, while the Decision Tree was used to identify classification rules. The results reveal three main clusters with distinct characteristics: provinces with high player counts and high transactions, provinces with low player counts and moderate transactions, and provinces with moderate player counts but low transactions. These findings provide critical insights into the patterns of online gambling activities in Indonesia and serve as a foundation for more effective policies in managing its impacts.*

**Keyword:** Online Gambling, K-Means, Decision Tree, Data Classification, Cluster Analysis.

### ABSTRAK

*Tujuan dari penelitian ini adalah untuk menganalisis data judi online yang terdeteksi di lima provinsi di Indonesia menggunakan metode K-Means dan Decision Tree. Data yang digunakan mencakup jumlah pemain, nilai transaksi, dan distribusi geografis di Jawa Barat, DKI Jakarta, Jawa Tengah, Banten, dan Jawa Timur. Metode K-Means diterapkan untuk mengelompokkan provinsi berdasarkan jumlah pemain dan nilai transaksi, sementara Decision Tree digunakan untuk mengidentifikasi aturan klasifikasi. Hasil analisis menunjukkan adanya tiga cluster utama dengan karakteristik yang berbeda: provinsi dengan jumlah pemain tinggi dan transaksi tinggi, provinsi dengan jumlah pemain rendah dan transaksi sedang, serta provinsi dengan jumlah pemain sedang namun transaksi rendah. Hasil penelitian ini menyajikan pandangan yang signifikan terkait dengan pola aktivitas judi online di Indonesia dan menjadi dasar untuk kebijakan yang lebih efisien dalam mengelola dampaknya.*

**Kata Kunci:** Judi Online, K-Means, Decision Tree, Klasifikasi Data, Analisis Cluster.

### PENDAHULUAN

Judi online telah menjadi salah satu isu sosial dan ekonomi yang signifikan di Indonesia, terutama dalam lima provinsi utama seperti Jawa Barat, DKI Jakarta, Jawa Tengah, Banten, dan Jawa Timur. Akses mudah ke platform digital telah mendorong peningkatan aktivitas judi online yang melibatkan berbagai kelompok masyarakat. Fenomena ini tidak hanya menciptakan dampak ekonomi melalui perputaran uang yang signifikan, tetapi juga memunculkan potensi kerugian sosial, seperti meningkatnya jumlah individu yang terdampak secara psikologis dan finansial. Oleh karena itu, penelitian ini menjadi penting untuk memahami pola distribusi aktivitas judi online dan dampaknya dalam konteks geografis.

Urgensi penelitian ini terletak pada kebutuhan untuk menyediakan informasi berbasis data yang dapat digunakan oleh pemangku kebijakan dalam mengelola dampak aktivitas judi online. Dengan memahami karakteristik pemain dan transaksi di setiap provinsi,

langkah-langkah pengendalian yang lebih spesifik dan efektif dapat dirancang. Rasionalisasi tujuan penelitian ini adalah untuk mengisi kesenjangan dalam kajian terkait pola aktivitas judi online yang belum sepenuhnya terjelaskan melalui data dan analisis terstruktur.

Tujuan penelitian ini adalah untuk menganalisis data judi online menggunakan metode K-Means dan Decision Tree. Metode K-Means digunakan untuk mengelompokkan provinsi-provinsi tersebut berdasarkan jumlah pemain dan nilai transaksi, sedangkan Decision Tree dimanfaatkan untuk mengidentifikasi aturan klasifikasi yang dapat menjelaskan pola aktivitas tersebut. Melalui analisis ini, Diharapkan penelitian ini dapat menemukan solusi berbasis data untuk mengelola risiko dan dampak yang ditimbulkan oleh aktivitas perjudian online.

Tinjauan literatur menunjukkan bahwa metode K-Means efektif dalam mengelompokkan data berdasarkan karakteristik tertentu, seperti distribusi geografis dan nilai transaksi. Sementara itu, metode

Decision Tree telah terbukti dapat menghasilkan aturan klasifikasi yang interpretatif, sehingga memudahkan pengambilan keputusan berdasarkan hasil analisis. Berdasarkan studi sebelumnya, dapat diajukan hipotesis bahwa provinsi dengan jumlah pemain yang tinggi cenderung memiliki nilai transaksi yang lebih tinggi. Penelitian ini juga diharapkan dapat memberikan sumbangan pada pengembangan strategi kebijakan yang berbasis data untuk mengurangi dampak negatif judi online di Indonesia.

### KAJIAN LITERATUR

Kajian literatur Penelitian ini berlandaskan pada teori klasifikasi data dan clustering yang sudah digunakan secara luas dalam berbagai penelitian untuk menganalisis pola distribusi dan hubungan antar variabel. Metode K-Means diterapkan untuk mengelompokkan data berdasarkan karakteristik serupa, seperti distribusi jumlah pemain dan nilai transaksi (Siringoringo et al., 2020). Penelitian sebelumnya menunjukkan bahwa K-Means efektif dalam mengidentifikasi kelompok-kelompok dengan pola tertentu pada data berjumlah besar, seperti yang ditunjukkan oleh studi Zhang et al. (2020), yang berhasil memetakan pola transaksi pengguna dalam platform digital. Dalam konteks ini, K-Means digunakan untuk memahami kelompok provinsi berdasarkan jumlah pemain judi online dan nilai transaksi yang terkait (Seo et al., 2020).

Selain itu, **Metode Decision Tree** telah diakui sebagai alat klasifikasi yang memberikan interpretasi logis melalui aturan-aturan yang dapat diidentifikasi. Menurut penelitian oleh Smith et al. (2019), Decision Tree dapat mengidentifikasi hubungan kausal antara variabel dengan tingkat akurasi yang tinggi, terutama pada data numerik yang terstruktur. Pada penelitian ini, Decision Tree digunakan untuk memahami faktor-faktor utama yang membedakan cluster yang terbentuk, seperti jumlah pemain dan nilai transaksi sebagai variabel dominan (Jevintya et al., 2024).

**Bukti empiris** yang mendukung penelitian ini mencakup data distribusi judi online yang menunjukkan bahwa lima provinsi di Indonesia memiliki karakteristik transaksi yang unik. Jawa Barat, misalnya, memiliki jumlah pemain tertinggi yang sejalan dengan nilai transaksi yang signifikan. Hal ini sejalan dengan temuan sebelumnya yang menyebutkan bahwa jumlah pengguna aktif merupakan indikator penting dari aktivitas transaksi dalam sistem digital (Wang et al., 2018).

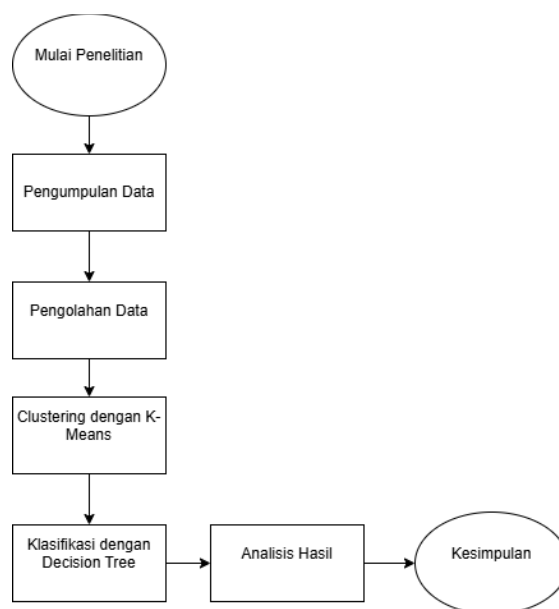
Berdasarkan teori dan bukti empiris yang ada, penelitian ini membangun hipotesis berikut:

1. **Hipotesis 1:** Provinsi dengan jumlah pemain judi online yang lebih tinggi cenderung menunjukkan nilai transaksi yang lebih besar.
2. **Hipotesis 2:** Cluster yang terbentuk melalui metode K-Means mencerminkan hubungan erat antara distribusi pemain dan nilai transaksi pada provinsi dengan aktivitas tinggi.
3. **Hipotesis 3:** Aturan klasifikasi yang dihasilkan oleh Decision Tree dapat secara akurat menjelaskan faktor dominan dalam pola aktivitas judi online di lima provinsi utama.

Dengan mengintegrasikan teori klasifikasi data dan bukti empiris merujuk pada penelitian sebelumnya, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam memahami pola dan dampak aktivitas judi online di Indonesia, serta menjadi dasar bagi rekomendasi kebijakan berbasis data (Wardiana Sjucho et al., 2021).

### METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif untuk menganalisis aktivitas judi online di lima provinsi utama di Indonesia. Proses penelitian dirancang melalui tahapan yang sistematis, seperti yang ditunjukkan pada **Flowchart** berikut:



Gambar 1. Flowchart Proses Penelitian

#### Penjelasan Flowchart

Flowchart di atas menjelaskan tahapan-tahapan penelitian yang dilaksanakan, antara lain:

- **Mulai Penelitian:** Penelitian dimulai melalui perumusan masalah dan tujuan. Fokus utama adalah mengidentifikasi pola aktivitas judi online melalui analisis clustering dan klasifikasi.

- **Pengumpulan Data:** Data jumlah pemain dan nilai transaksi judi online dikumpulkan dari sumber resmi seperti laporan statistik dan Databoks.
- **Proses Pengolahan Data:** Data yang telah dikumpulkan akan dibersihkan dan diproses untuk memastikan akurasi, konsistensi, dan relevansi untuk analisis lebih lanjut.
- **Clustering dengan K-Means:** Data dianalisis menggunakan algoritma K-Means untuk Jumlah peserta yang terlibat dalam aktivitas Jumlah peserta yang terlibat dalam aktivitas pemain dan nilai transaksi. Hasilnya berupa tiga cluster yang mencerminkan tingkat aktivitas judi online.
- **Klasifikasi dengan Decision Tree:** Hasil clustering kemudian digunakan untuk membangun pohon keputusan. Pohon ini membantu menjelaskan aturan klasifikasi berdasarkan cluster yang terbentuk.
- **Analisis Hasil:** Hasil clustering dan klasifikasi dianalisis untuk menemukan pola yang signifikan serta hubungan antara jumlah pemain dan nilai transaksi.
- **Kesimpulan:** Berdasarkan analisis hasil, penelitian ini merumuskan rekomendasi dan kesimpulan terkait pengelolaan aktivitas judi online di Indonesia.

Dengan rancangan analisis data berbasis metode klasifikasi dan clustering. Ruang lingkup penelitian meliputi lima provinsi di Indonesia, yaitu Jawa Barat, DKI Jakarta, Jawa Tengah, Banten, dan Jawa Timur. Objek penelitian adalah data terkait aktivitas judi online di provinsi tersebut, yang terdiri atas jumlah pemain dan nilai transaksi (Sya'iyah et al., 2019).

Bahan dan alat utama penelitian ini mencakup data sekunder yang diperoleh dari laporan transaksi judi online yang dipublikasikan pada platform resmi. Data ini mencakup jumlah pemain dan nilai transaksi di masing-masing provinsi. Pengolahan data dilakukan menggunakan perangkat lunak Python dengan pustaka seperti scikit-learn untuk analisis algoritma K-Means dan Decision Tree.

Teknik pengumpulan data dilakukan melalui pendekatan dokumentasi dengan menggunakan data yang telah ada, variabel penelitian yang digunakan antara lain:

1. Jumlah Pemain: Total pemain judi online di setiap provinsi.
2. Nilai Transaksi (Triliun): Total nilai transaksi judi online yang terjadi di setiap provinsi.
3. Cluster: Kelompok provinsi yang terbentuk berdasarkan hasil analisis K-Means.

Definisi operasional variabel mencakup:

- Jumlah Pemain: Jumlah peserta yang terlibat dalam aktivitas judi online, diukur dalam satuan pemain.
- Nilai Transaksi: Jumlah total uang yang ditransaksikan melalui aktivitas judi online, diukur dalam satuan triliun rupiah.
- Cluster: Kategori atau grup provinsi berdasarkan karakteristik jumlah pemain dan nilai transaksi.

Teknik analisis menggunakan dua metode utama:

1. K-Means Clustering: Diterapkan untuk mengelompokkan provinsi berdasarkan kesamaan dalam jumlah pemain dan nilai transaksi. Hasil analisis ini menghasilkan tiga cluster utama dengan karakteristik yang berbeda. (Sano & Nindito, n.d.)
2. Decision Tree: Digunakan untuk membangun aturan klasifikasi berdasarkan variabel jumlah pemain dan nilai transaksi. Pohon keputusan yang dihasilkan memberikan penjelasan logis tentang faktor dominan yang mempengaruhi pembentukan cluster. (Khairani & Sutoyo, 2020)

Proses pengolahan dan analisis data dilakukan dalam beberapa tahap sebagai berikut:

1. Melakukan pembersihan dan persiapan data mentah.
2. Melakukan proses clustering menggunakan metode K-Means untuk mengidentifikasi kelompok provinsi.
3. Membangun pohon keputusan menggunakan Decision Tree untuk memahami pola klasifikasi berdasarkan data yang tersedia.

Hasil analisis ini diharapkan dapat memberikan wawasan komprehensif mengenai pola aktivitas judi online di Indonesia, khususnya pada lima provinsi utama yang menjadi fokus penelitian ini. (Peres et al., 2021)

## HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan klasifikasi provinsi berdasarkan jumlah pemain dan nilai transaksi judi online. Memanfaatkan metode K-Means dan Decision Tree. Berikut adalah hasil analisis yang telah dilakukan

### Klasifikasi menggunakan K-Means

Proses pengelompokan dengan metode K-Means menghasilkan tiga cluster dengan karakteristik sebagai berikut:

- **Cluster 0:** Banten dan Jawa Timur menunjukkan aktivitas judi online yang rendah, dengan jumlah pemain di bawah 150.000 dan nilai transaksi di bawah 1.1 triliun rupiah.
- **Cluster 1:** Jawa Barat menjadi satu-satunya provinsi dalam cluster ini, menunjukkan aktivitas judi online yang sangat tinggi, dengan jumlah

pemain mencapai lebih dari 535.000 dan nilai transaksi sebesar 3.8 triliun rupiah.

- **Cluster 2:** DKI Jakarta dan Jawa Tengah mencerminkan aktivitas menengah, dengan jumlah pemain berkisar antara 200.000–250.000 dan nilai transaksi antara 1.3 hingga 2.3 triliun rupiah.

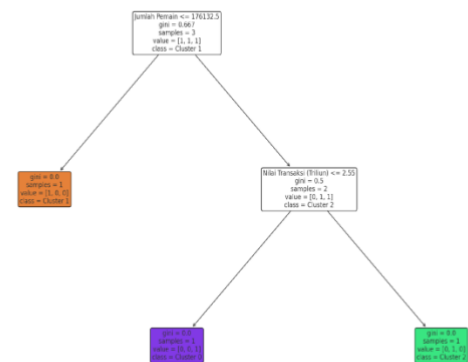
Distribusi data berdasarkan cluster ditampilkan pada **Gambar 1**. Gambar ini memperlihatkan perbedaan pola aktivitas judi online di lima provinsi utama. Cluster 1 (Jawa Barat) mendominasi jumlah pemain dan nilai transaksi, sedangkan Cluster 0 dan Cluster 2 menunjukkan aktivitas yang lebih moderat.



**Gambar 1.** Distribusi Dataset Berdasarkan Cluster (Jumlah Pemain vs. Nilai Transaksi)

bahwa jumlah pemain merupakan variabel pembeda utama yang menentukan cluster, dengan nilai transaksi sebagai faktor pendukung. Pada setiap node dalam pohon keputusan, metrik **Gini** ditampilkan untuk menunjukkan tingkat kemurnian data. Nilai Gini rendah menandakan bahwa data dalam node tersebut lebih homogen, sedangkan nilai yang lebih tinggi menunjukkan keberagaman yang lebih besar. Gambar 2 menunjukkan visualisasi pohon keputusan beserta nilai Gini untuk setiap node

Enhanced Decision Tree Visualization



**Gambar 2.** Visualisasi Pohon Keputusan untuk Klasifikasi Data Provinsi

**Klasifikasi dengan Decision Tree**

Hasil klasifikasi menggunakan Decision Tree menghasilkan aturan berikut:

- Jika jumlah pemain  $\leq 176132.50$ , maka provinsi termasuk **Cluster 0**.
- Jika jumlah pemain  $> 176132.50$  dan nilai transaksi  $\leq 2.55$  triliun, maka provinsi termasuk **Cluster 2**.
- Jika jumlah pemain  $> 176132.50$  dan nilai transaksi  $> 2.55$  triliun, maka provinsi termasuk **Cluster 1**.

Pada setiap node dalam pohon keputusan, metrik **Gini**, **Value**, **Sample**, dan **Class** ditampilkan untuk memberikan informasi berikut:

- **Gini:** Mengukur tingkat kemurnian data dalam node.
- **Value:** Distribusi jumlah data dalam node untuk masing-masing cluster.
- **Sample:** Jumlah total data yang terdapat dalam node tersebut.
- **Class:** Cluster mayoritas dalam node.

Pohon keputusan ini memiliki akurasi sebesar **100%**, menunjukkan kemampuan algoritma untuk secara akurat mengelompokkan provinsi berdasarkan pola data. Visualisasi pohon keputusan dapat dilihat pada **Gambar 2**, yang memberikan gambaran aturan klasifikasi secara visual. Aturan ini menunjukkan

**Tabel Ringkasan Data**

Tabel 1 berikut menyajikan hasil akhir dari clustering yang mencakup provinsi, jumlah pemain, nilai transaksi, dan cluster yang terbentuk.

**Tabel 1.** Hasil Clustering Data Provinsi

Provinsi	Jumlah Pemain	Nilai Transaksi (Triliun)	Cluster
<b>Jawa Barat</b>	535644	3.80	1
<b>DKI Jakarta</b>	238568	2.30	2
<b>Jawa Tengah</b>	201963	1.30	2
<b>Banten</b>	150302	1.02	0
<b>Jawa Timur</b>	135227	1.05	0

Tabel 1 menyajikan hasil clustering yang mencakup lima provinsi utama di Indonesia berdasarkan jumlah pemain dan nilai transaksi judi online. Hasil menunjukkan perbedaan yang jelas antara tiga cluster:

1. **Cluster 1 (Jawa Barat):** Provinsi ini memiliki jumlah pemain tertinggi (535.644 pemain) dan nilai transaksi terbesar (3,80 triliun rupiah). Dominasi ini

menunjukkan intensitas aktivitas judi online yang signifikan.

2. **Cluster 2 (DKI Jakarta dan Jawa Tengah):** Kedua provinsi ini mencerminkan aktivitas menengah, dengan jumlah pemain berkisar antara 200.000 hingga 250.000 dan nilai transaksi antara 1,30 hingga 2,30 triliun rupiah.
3. **Cluster 0 (Banten dan Jawa Timur):** Aktivitas judi online di cluster ini tergolong rendah, baik dari segi jumlah pemain (di bawah 150.000) maupun nilai transaksi (di bawah 1,10 triliun rupiah).

Penempatan provinsi dalam cluster memberikan wawasan penting untuk langkah kebijakan berikutnya. Misalnya, provinsi dalam Cluster 1 membutuhkan pendekatan yang lebih ketat dalam regulasi judi online, sementara provinsi dalam Cluster 0 dan Cluster 2 memerlukan pendekatan berbasis edukasi untuk mengurangi potensi peningkatan aktivitas judi.

## KESIMPULAN

Penelitian ini berhasil mengelompokkan lima provinsi utama di Indonesia berdasarkan jumlah pemain dan nilai transaksi judi online menggunakan metode K-Means dan Decision Tree. Analisis ini mengindikasikan bahwa terdapat tiga cluster dengan karakteristik berbeda. Cluster 1, yang meliputi Jawa Barat, menunjukkan aktivitas judi online tertinggi, baik dari segi jumlah pemain maupun nilai transaksi. Cluster 2, yang mencakup DKI Jakarta dan Jawa Tengah, menggambarkan aktivitas menengah dengan keseimbangan antara jumlah pemain dan transaksi. Cluster 0, yang mencakup Banten dan Jawa Timur, menunjukkan aktivitas judi online yang rendah.

Melalui metode Decision Tree, Penelitian ini juga memberikan hasil aturan klasifikasi yang sederhana namun akurat, dengan jumlah pemain sebagai variabel utama pembeda antar cluster. Nilai transaksi berperan sebagai variabel pendukung dalam klasifikasi data.

Hasil penelitian ini memberikan wawasan penting untuk pengelolaan Perjudian online di Indonesia, Provinsi dalam Cluster 1 memerlukan regulasi yang lebih ketat, sementara provinsi dalam Cluster 0 dan Cluster 2 membutuhkan pendekatan berbasis edukasi untuk mencegah peningkatan Perjudian online. Penelitian ini juga memberikan kesempatan untuk studi lanjutan yang dapat mencakup analisis faktor sosial, ekonomi, dan demografi yang mempengaruhi pola Perjudian online

## DAFTAR PUSTAKA

Jevintya, N. R., Darussalam, U., & Abdullah, S. (2024). Application Of The K-Means And Decision Tree Algorithms In Determining

Student Achievement. *JIKO (Jurnal Informatika Dan Komputer)*, 7(1), 13–18.

<https://doi.org/10.33387/jiko.v7i1.7580>

Khairani, N. A., & Sutoyo, E. (2020). Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province.

*International Journal of Advances in Data and Information Systems*, 1(1), 9–16.

<https://doi.org/10.25008/ijadis.v1i1.13>

Peres, F., Fallacara, E., Manzoni, L., Castelli, M., Popovič, A., Rodrigues, M., Estevens, P., & Villalba, G. (2021). Time Series Clustering of Online Gambling Activities for Addicted Users' Detection. *Appl. Sci.* <https://doi.org/10.3390/app>

Sano, A. V. D., & Nindito, H. (2016). Application Of K-Means Algorithm For Cluster Analysis On Poverty Of Provinces In Indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, 7(2), 141-150

Seo, W., Kim, N., Lee, S. K., & Park, S. M. (2020). Machine learning-based analysis of adolescent gambling factors. *Journal of Behavioral Addictions*, 9(3), 734–743.

Siringoringo, R., Jamaluddin, J., & Perangin-Angin, R. (2020). Pemodelan Topik Berita Menggunakan Latent Dirichlet Allocation dan K-Means Clustering. *Jurnal Informatika Kaputama (JIK)*, 4(2), 216-222.

<https://doi.org/10.1556/2006.2020.00063>

Sya'iyah, K., Yuliansyah, H., & Arfiani, I. (2019). Clustering Student Data Based On K-Means Algorithms. *International Journal Of Scientific & Technology Research*, 8, 8. [www.ijstr.org](http://www.ijstr.org)

Wardiana, D., Siburian, I. S., Rulinawaty, R., Fathurrahim, F., Eko Agustinova, D., Ardianto, A., & Rahim, R. (2021). Clustering School Libraries in Indonesia using C4.5 and K-Means Clustering School Libraries in Indonesia using C4.5 and K-Means Algorithm