



## Analisis Jam Belajar dan Hasil Ujian Siswa dengan Menggunakan Metode K-Means Clustering, Principal Component Analysis (PCA)

Andreas Sijabat<sup>1\*</sup>, Satna Siagian<sup>2</sup>, Sepania Yolanda Br Sinaga<sup>3</sup>, Muhammad Yasin Naufal<sup>4</sup>

<sup>1-4</sup>Program Studi Informatika, Fakultas Teknologi dan Ilmu Komputer, Universitas Satya Terra Bhinneka, Indonesia

Korespondensi penulis: [andreassijabat772@gmail.com](mailto:andreassijabat772@gmail.com)

**Abstract.** Education is a fundamental pillar in shaping students' abilities and potential, with one of the critical factors often under attention being the time spent on studying. Understanding the extent to which study time impacts academic achievement is essential for improving education quality. This study aims to explore whether there is a significant relationship between the number of study hours per week and students' final exam scores. The methods used in this research include K-Means Clustering, Principal Component Analysis (PCA), and the Apriori Algorithm. The expected result of this study is to discover a significant correlation between the two variables. If a significant positive correlation is found, it would indicate that increased study time is associated with better exam performance. Conversely, if the correlation is weak or negative, it would suggest that other factors may play a more dominant role in influencing students' academic performance. Furthermore, significance testing is conducted to ensure that the results of the analysis did not occur by chance. This research is expected to provide valuable insights for parents and students in developing more effective and efficient learning strategies. For educators, these findings can serve as a foundation for designing more structured learning programs, including recommendations for optimal study time allocation. On the other hand, for students, the results of this study can serve as a guide to managing their study time more optimally.

**Keywords:** Correlation, Data Analysis, Education, Exam Scores, Study Hours.

**Abstrak.** Pendidikan merupakan pilar utama dalam membentuk kemampuan dan potensi siswa di mana salah satu faktor penting yang sering menjadi perhatian adalah waktu yang dihabiskan untuk belajar. Pemahaman mengenai seberapa besar pengaruh waktu belajar terhadap pencapaian akademik sangatlah penting untuk meningkatkan kualitas pendidikan. Penelitian ini bertujuan untuk mengeksplorasi apakah Terdapat hubungan yang signifikan antara jumlah jam belajar per minggu yang dihabiskan oleh siswa dengan nilai ujian akhir yang mereka peroleh. Metode yang digunakan dalam penelitian ini K-Means Clustering, Principal Component Analysis (PCA), Apriori Algorithm. Hasil yang diharapkan dari penelitian ini adalah penemuan adanya korelasi yang signifikan antara kedua variabel tersebut jika ditemukan korelasi positif yang signifikan artinya peningkatan waktu belajar dapat dikaitkan dengan peningkatan nilai ujian sebaliknya jika korelasi Lemah atau negatif maka hasil ini mengidentifikasi adanya faktor-faktor lain yang lebih dominan dalam mempengaruhi pencapaian akademik siswa. Selain itu uji signifikan dilakukan untuk memastikan bahwa hasil analisis ini tidak terjadi secara kebetulan. Penelitian ini diharapkan dapat memberikan wawasan penting bagi orang tua dan siswa dalam mengembangkan strategi belajar yang lebih efektif dan efisien. Bagi para pendidik temuan ini dapat digunakan sebagai dasar untuk merancang program pembelajaran yang lebih terstruktur termasuk rekomendasi alokasi waktu belajar yang optimal. Di sisi lain dari siswa hasil penelitian ini dapat menjadi panduan untuk mengatur waktu belajar mereka secara lebih optimal.

**Kata Kunci:** Analisis Data, Jam Belajar, Korelasi, Nilai Ujian, Pendidikan,

### 1. PENDAHULUAN

Pendidikan memiliki peran sentral dalam membentuk kualitas sumber daya manusia, terutama dalam mencetak generasi yang cakap secara akademis maupun sosial. Di tengah meningkatnya tuntutan terhadap pencapaian akademik siswa, salah satu aspek yang kerap menjadi perhatian adalah bagaimana pola dan durasi belajar mempengaruhi hasil belajar siswa. Jam belajar sering diasumsikan sebagai indikator langsung dari tingkat keseriusan dan upaya

siswa dalam memahami materi pelajaran. Namun, Apakah durasi belajar yang lebih lama selalu berbanding lurus dengan nilai ujian yang lebih tinggi masih menjadi pertanyaan yang relevan untuk diteliti lebih lanjut.

Dalam konteks ini, penerapan teknik data mining dan machine learning menjadi sangat relevan. Metode K-Means Clustering dapat digunakan untuk mengelompokkan siswa berdasarkan pola jam belajar dan hasil ujian mereka. Selanjutnya, Principal Component analysis (PCA) digunakan untuk mereduksi dimensi data agar pola-pola utama lebih mudah dikenali tanpa menghilangkan informasi penting.

Dalam konteks tersebut, penting dilakukan sebuah analisis yang komprehensif untuk memahami hubungan antara jam belajar dan hasil ujian siswa. Dengan kemajuan dalam bidang ilmu data, pendekatan berbasis analisis data kini menjadi salah satu metode yang efektif dalam mengevaluasi berbagai faktor yang mempengaruhi capaian akademik. Penelitian ini tidak hanya bertujuan untuk menjawab pertanyaan mengenai hubungan antara jam belajar dan nilai ujian, tetapi juga untuk memberikan pemahaman berbasis data pada pendidik, siswa, dan orang tua dalam merancang strategi belajar yang lebih efektif dan efisien. Dengan demikian, hasil dari studi ini diharapkan dapat menjadi referensi dalam pengambilan keputusan dunia pendidikan, khususnya terkait perancangan waktu belajar yang optimal untuk mendukung prestasi akademik siswa.

Melalui pendekatan ini, diadakan analisis yang dihasilkan tidak hanya bersifat deskriptif, tetapi juga maupun memberikan wawasan yang berguna dalam demikian, pernyataan ini memiliki urgensi dan kontribusi peneliti penting dalam upaya meningkatkan kualitas pembelajaran berbasis data.

## **2. TINJAUAN PUSTAKA**

### **Jam Belajar**

Jam belajar adalah waktu yang dialokasikan oleh siswa untuk melakukan kegiatan belajar secara mandiri maupun terstruktur di luar kegiatan pembelajaran formal. Durasi belajar ini diyakini memiliki pengaruh terhadap pencapaian akademik siswa. Menurut Slameto (2010), keberhasilan belajar tidak hanya ditentukan oleh kecerdasan tetapi juga oleh usaha belajar yang dilakukan, termasuk durasi dan intensitasnya.

### **Nilai Ujian**

Nilai ujian merupakan indikator formal yang digunakan untuk mengukur pencapaian kompetensi Akademi siswa. Nilai ini dapat dipengaruhi oleh berbagai faktor, seperti kemampuan kognitif, metode belajar, motivasi, lingkungan belajar, dan durasi belajar. Oleh

karena itu, penting untuk menganalisis hubungan antara variabel dan belajar dan nilai ujian secara statistik dan komputasional.

### **Data Mining dalam Dunia Pendidikan**

Data mining adalah proses menemukan pola atau informasi yang bermanfaat dari kumpulan data besar dengan menggunakan metode statistik, matematika, dan kecerdasan buatan. Dalam konteks pendidikan, data mining digunakan untuk menganalisis perilaku belajar siswa, memprediksi prestasi akademik, dan memberikan rekomendasi pembelajaran.

#### ***K-Means Clustering***

K-Means Clustering adalah algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam jumlah *cluster* berdasarkan kemiripan. Dalam konteks penelitian ini, metode ini digunakan untuk mengelompokkan siswa berdasarkan pola jam belajar dan nilai ujian yang dimiliki. Tujuannya adalah untuk mengidentifikasi kelompok-kelompok siswa dengan karakteristik tertentu seperti rajin namun nilai rendah atau belajar sedikit namun hasil tinggi.

#### **Principal Component Analysis (PCA)**

PCA merupakan teknik reduksi dimensi yang bertujuan untuk menyederhanakan data berdimensi tinggi menjadi dimensi yang lebih rendah dengan tetap mempertahankan informasi penting. Dalam penelitian ini, PCA digunakan untuk memudahkan visualisasi dan pemahaman terhadap variabel-variabel yang paling mempengaruhi hasil ujian siswa.

### **3. METODOLOGI PENELITIAN**

#### **Jenis Penelitian**

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan eksperimental berbasis sistem aplikasi interaktif. Penelitian ini menggunakan implementasi langsung algoritma K-Means Clustering, Principal Component Analysis (PCA) dalam bentuk sistem berbasis web yang dikembangkan menggunakan JavaScript.

#### **Objek dan Subjek Penelitian**

Objek dari penelitian ini adalah data mahasiswa yang meliputi nama, jenis kelamin, usia, semester, jam belajar per hari, dan nilai ujian. Subjek penelitian adalah mahasiswa perguruan tinggi yang mengisi formulir digital dari google form.

#### **Instrumen Penelitian**

Instrumen penelitian berupa sistem web interaktif yang berfungsi untuk:

- 1) Mengumpulkan data mahasiswa (form input),
- 2) Menganalisis data menggunakan K-Means, PCA.

- 3) Menampilkan hasil dalam bentuk tabel dan grafik secara otomatis.

### **Teknik Pengumpulan Data**

Data dikumpulkan melalui formulir digital yang diisi langsung oleh mahasiswa pada google form. Data yang dikumpulkan mencakup:

- 1) Nama dan jenis kelamin,
- 2) Usia,
- 3) Semester,
- 4) Jam belajar per hari,
- 5) Nilai ujian.

### **Teknik Analisis Data**

Berikut adalah langkah-langkah analisis data:

#### **1) Analisis K-Means Clustering**

Mahasiswa dikelompokkan berdasarkan nilai PCA (gabungan antara jam belajar dan nilai) terhadap rata-rata. Proses ini membagi mahasiswa ke dalam beberapa klaster seperti klaster performa tinggi, sedang, dan rendah.

#### **2) Analisis PCA (Principal Component Analysis)**

PCA dihitung dengan rumus:

$$PCA = \text{Jam Belajar} + (\text{Nilai} \times 5) / 2$$

Rumus ini menyederhanakan dimensi jam belajar dan nilai ujian ke dalam satu skor gabungan yang dapat divisualisasikan dan diklasifikasikan lebih lanjut.

## **4. HASIL DAN PEMBAHASAN**

### **Deskripsi Sistem dan Data**

Sistem dikembangkan dalam bentuk aplikasi web interaktif yang memproses data mahasiswa secara langsung. Setelah data diinput oleh pengguna, sistem akan menampilkan:

- 1) Jumlah siswa terdaftar,
- 2) Rata-rata nilai mahasiswa,
- 3) Jumlah siswa berperforma tinggi,
- 4) Distribusi nilai dan jam belajar dalam bentuk grafik.

### **Hasil K-Means Clustering**

Hasil klasterisasi dari sistem menunjukkan pembagian mahasiswa ke dalam tiga klaster berdasarkan nilai PCA:

Cluster 1: Jam belajar dan nilai tinggi,

Cluster 2: Jam belajar sedang, nilai sedang,

Cluster 3: Jam belajar rendah, nilai rendah.

Pembagian ini membantu dalam mengidentifikasi karakteristik mahasiswa dan merancang strategi belajar yang sesuai berdasarkan kelompoknya.

### **Hasil Principal Component Analysis (PCA)**

PCA menyajikan informasi dalam dimensi tunggal yang memadukan dua variabel utama. Visualisasi PCA dalam bentuk diagram batang atau sebar menunjukkan kecenderungan hubungan antara kombinasi belajar dan pencapaian nilai.

Contoh:

Mahasiswa dengan jam belajar 4 jam/hari dan nilai 90 akan memiliki nilai  $PCA = (4 + (90/100 * 5))/2 = 4.25$

Nilai PCA ini kemudian dibandingkan dengan rata-rata untuk menentukan klaster siswa tersebut.

### **Visualisasi dan Tampilan Sistem**

Sistem juga menyediakan berbagai tampilan visualisasi seperti:

Grafik distribusi klaster siswa,

Grafik PCA,

Grafik distribusi nilai,

Grafik hubungan jam belajar vs nilai.

Visualisasi ini memudahkan pemahaman data secara menyeluruh dan mendukung pengambilan keputusan berbasis bukti.

### **Pembahasan**

Hasil yang diperoleh dari sistem menunjukkan bahwa penggunaan algoritma machine learning sederhana mampu memberikan insight yang berguna tentang hubungan antara jam belajar dan nilai. Tidak semua siswa yang belajar lama memperoleh nilai tinggi, dan efisiensi belajar serta waktu belajar menjadi faktor penting yang terungkap melalui analisis ini.

## **5. KESIMPULAN DAN SARAN**

Berdasarkan hasil analisis yang telah dilakukan menggunakan metode K-Means Clustering dan Principal Component Analysis (PCA), dapat disimpulkan bahwa terdapat korelasi positif antara jam belajar dan nilai ujian siswa, meskipun hubungan tersebut tidak selalu linier. Beberapa siswa dengan jam belajar tinggi belum tentu memperoleh nilai yang tinggi, yang menunjukkan bahwa kualitas dan strategi belajar juga memegang peranan penting. Metode K-Means Clustering berhasil mengelompokkan siswa ke dalam tiga klaster utama, yaitu: (a) siswa dengan jam belajar dan nilai tinggi, (b) siswa dengan jam belajar dan nilai

sedang, serta (c) siswa dengan jam belajar dan nilai rendah. Hal ini mengindikasikan adanya segmentasi performa yang dapat dijadikan dasar dalam merancang rekomendasi pembelajaran yang lebih personal. PCA terbukti efektif dalam menyederhanakan data, memungkinkan sistem untuk menampilkan skor gabungan antara jam belajar dan nilai siswa sehingga mempermudah visualisasi dan penentuan kluster. Selain itu, penggunaan aplikasi berbasis web dalam analisis ini turut mempermudah proses pengolahan, pengelompokan, dan interpretasi data secara otomatis serta interaktif.

Berdasarkan temuan tersebut, disarankan bagi siswa untuk tidak hanya menambah durasi belajar, tetapi juga meningkatkan kualitas dan efisiensi waktu belajar agar hasil akademik yang diperoleh lebih optimal. Bagi para pendidik dan pengambil kebijakan pendidikan, hasil ini dapat dimanfaatkan untuk merancang strategi pembelajaran berbasis data, termasuk memberikan perhatian khusus pada kelompok siswa dengan performa rendah. Selain itu, pengembangan sistem lebih lanjut dapat dilakukan dengan menambahkan fitur prediksi nilai ujian berdasarkan histori belajar siswa, serta mempertimbangkan variabel tambahan seperti jenis materi yang dipelajari, metode pembelajaran yang digunakan, dan kondisi psikologis siswa guna meningkatkan akurasi dan relevansi hasil analisis.

## **DAFTAR PUSTAKA**

- Ahmad Pandhu Wijaya, & Fachreza, A. (2024, Maret). Pengelompokan hasil belajar siswa dengan metode clustering K-Means pada SMK Muhammadiyah 1 Weleri. *Estudiar*.
- Berlian Juliartha Martin Putra, & Yuniarti, D. A. F. (n.d.). Prediksi nilai akhir mata kuliah mahasiswa menggunakan metode K-Means Clustering. *Teknosi – Universitas Andalas*.
- Borgavakar, S. P., & Shrivastava, A. (2018). Evaluating student's performance using K-Means Clustering. *International Journal of Engineering Research & Technology (IJERT)*.
- Borgavakar, S. P., & Shrivastava, A. (2022). Data analysis of educational evaluation using K-Means Clustering. *PMC*.
- Dewi, S., Defit, S., & Yuhandri, Y. (2023). Akurasi pemetaan kelompok belajar siswa menuju prestasi menggunakan metode K-Means. *Jurnal Sistem Informasi dan Teknologi*.
- Gajewski, Ł., Chołoniowski, J., & Hołyst, J. (2016, April). Key courses of academic curriculum uncovered by data mining of students' grades. *arXiv*. <https://arxiv.org/abs/1604.04265>
- IamOn, N., & Boongoen, T. (2018). Generating descriptive model for student dropout: A review of clustering approach.

- Islam, M. H. S., & Haque, M. (2012). An approach of improving students' academic performance by using K-Means Clustering and Decision Tree. arXiv preprint. <https://arxiv.org/abs/1203.3376>
- Li, R. (2023). Predicting student academic performance. Semantic Scholar.
- Putri, T. N., Renita, C., & Wahyu, M. A. E. (2024). Identifikasi pola belajar dan absensi siswa dengan teknik clustering dan analisis korelasi. *Jurnal JATI*, 8(5).
- Sri Dewi, Defit, S., & Yuhandri, Y. (2023). Akurasi pemetaan kelompok belajar siswa menuju prestasi menggunakan metode K-Means. *Jurnal Sistem Informasi dan Teknologi*.
- Tyas Widyanti, Hilabi, S. S., Hananto, A., Tukino, & Novalia, E. (2023). Implementasi K-Means dan K-Nearest Neighbors pada kategori siswa berprestasi. *Jurnal Informasi dan Teknologi*.
- Wijaya, Y., Yudhistira, A., & Andika, R. (2023, Maret). Pengelompokan data nilai siswa menggunakan metode K-Means Clustering. *JAITI*, 1(1). E-Journal Poliban.
- Yuniarti, D. A. F., & Martin Putra, B. J. (n.d.). Analisis hasil belajar mahasiswa dengan clustering menggunakan metode K-Means. *Poros Teknik*.