



Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Pendapatan Orang Tua Terhadap Hasil Belajar Pada Mata Pelajaran Produktif

Nanndo Yannuansa¹, Muhamad Safa'udin², Mochammad Ilham Aziz³

¹Teknik Elektro, Fakultas Teknik, Universitas Hasyim Asy'A'ari

²Pendidikan Guru Madrasah Ibtida'iyah, Fakultas Tarbiyah, Institut Agama Islam Tribakti Kediri

³Teknik Informatika, Fakultas Teknik, Universitas Dian Nuswantoro
Email: n4nnd0yan@gmail.com, safaudin7@gmail.com

Abstrak

Tingginya tingkat keberhasilan siswa dan rendahnya tingkat kegagalan siswa merupakan cemin kualitas dunia pendidikan. Dunia pendidikan saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya manusia dan sistem informasi, sarana dan prasarana merupakan salah satu sumber daya yang dapat digunakan untuk meningkatkan kemampuan bersaing. Penelitian ini bertujuan untuk mengetahui hasil clusterisasi pada data Hasil belajar siswa dengan pemanfaatan algoritma K-Means Clustering untuk pengambilan kebijakan strategi dalam proses pembelajaran. Untuk Algoritma K-Means Clustering, data dikelompokkan berdasarkan karakteristik yang sama dan dimasukkan kedalam kelompok atau cluster. Informasi yang akan ditampilkan berupa pengelompokan pengaruh hasil belajar siswa dan dibagi menjadi 3 kelompok. Hasil akhir dari penelitian ini bahwa dengan menggunakan Algoritma K-Means Clustering adalah kelompok siswa yang termasuk mendapatkan hasil kurang rata-rata orang tua mendapatkan penghasilan pas-pasan, dan siswa yang mendapatkan hasil baik orang tua siswa adalah golongan dari orang mampu karena untuk menunjang nilai produktif diperlukan alat-alat praktikum yang memadai dan siswa dapat menguasai terlebih kamera dan alat video recorder, sedangkan siswa yang masuk kedalam kelompok sedang orang tua berpenghasilan standar yaitu golongan menengah.

Kata Kunci : K-Means, Student, Data Mining

Abstract

The high level of student success and the low level of student failure is a reflection of the quality of the world of education. The world of education today is required to have the ability to compete by utilizing all available resources. In addition to human resources and information systems, facilities and infrastructure are one of the resources that can be used to improve competitiveness. This study aims to determine the results of clustering on student learning outcomes using the K-Means Clustering algorithm for strategic policy making in the learning process. For the K-Means Clustering Algorithm, the data are grouped based on the same characteristics and are included in groups or clusters. The information that will be displayed is in the form of grouping the influence of student learning outcomes and is divided into 3 groups. The final result of this study is that using the K-Means Clustering Algorithm is a group of students who get less than average results, parents earn mediocre income, and students who get good results, parents are a group of people who are able to support productive values. Adequate practical tools are needed and students can master cameras and video recorders, while students who fall into the middle group are parents with standard income, namely the middle class.

Keywords : K-Means, Student, Data Mining

A. Pendahuluan

Pendidikan merupakan salah satu cara suatu bangsa untuk terbebas dari ketertinggalan dan kebodohan. Salah satunya di Negara Indonesia, pendidikan merupakan suatu hal yang pokok. sesuai dengan Undang-Undang Dasar 1945 pasal 31 ayat (1) yang mengamatkan bahwa pemerintah mengusahakan dan menyelenggarakan satu sistem pendidikan nasional, yang meningkatkan keimanan dan ketakwaan serta akhlak mulia dalam rangka mencerdaskan bangsa dengan pendidikan. Pendidikan menurut pengertian secara psikologis merupakan suatu proses perubahan, yaitu perubahan tingkah laku sebagai hasil dari interaksi dengan lingkungannya dalam memenuhi kebutuhan hidupnya. Perubahan-perubahan tersebut akan nyata dalam seluruh aspek tingkah laku. Dan pengertian belajar adalah suatu perubahan tingkah laku seseorang untuk

memperoleh suatu perubahan tingkah laku yang secara keseluruhan, sebagai hasil pengalamannya sendiri dalam interaksi dengan lingkungannya.

Prestasi belajar dipengaruhi oleh faktor internal dan eksternal. Keluarga adalah salah satu faktor eksternal yang mempengaruhi prestasi belajar. Beberapa faktor diantaranya adalah tingkat pendidikan dan tingkat pendapatan orang tua. Tingkat pendidikan orang tua yang tinggi pasti memiliki cita-cita tinggi terhadap pendidikan anaknya. Sedangkan tingkat pendapatan orang tua dapat membantu siswa dalam memenuhi kebutuhan sekolahnya sehingga siswa dapat menjalani proses pendidikannya dengan maksimal. Keberhasilan pendidikan dapat dilihat melalui prestasi belajar yang diraih oleh siswa setelah melakukan proses pembelajaran. Prestasi belajar merupakan hasil dari usaha yang dicapai siswa selama melakukan kegiatan belajar di sekolah yang menghasilkan sebuah nilai.

Dalam makalah ini, peneliti mencoba memanfaatkan Algoritma K-Means Clustering untuk mengolah data hasil belajar siswa apakah pendapatan orang tua itu mempengaruhi terhadap hasil belajar siswa.

Algoritma pengelompokan K-means umumnya digunakan saat ini. Ini perlu terlebih dahulu menentukan jumlah kategori dan pusat pengelompokan awal yang sesuai, menetapkan setiap nilai ke kategori tertentu sesuai dengan prinsip jarak minimum, kemudian secara konstan menghitung setiap pusat kategori dan menyesuaikan kategori dari setiap fitur, dan mengambil jumlah minimum jarak kuadrat dari setiap fitur ke pusat kategori sebagai kriteria konvergensi akhir [10].

Literatur Review

Yufei Ding, Yue Zhao, Xipeng Shen, Madanlal Musuvathi, and Todd Mytkowicz (2015). Penelitian ini menunjukkan bahwa Yinyang K-means secara konsisten lebih cepat daripada algoritma sebelumnya, terlepas dari dimensi dan ukuran kumpulan data, jumlah cluster, dan konfigurasi mesin. Hal tersebut mempercepat tugas dan langkah pembaruan pusat di K-means, secara otomatis menghasilkan tradeoff yang baik antara biaya ruang dan peningkatan kinerja. Desainnya yang elastis membuatnya secara otomatis memaksimalkan kinerjanya di bawah batasan ruang tertentu. Ini mempertahankan semantik K-means asli. Sifat-sifat menarik ini, ditambah kesederhanaannya, menjadikannya sebagai pengganti praktis dari K-means standar selama pertidaksamaan segitiga berlaku[1].

Jian Wang, Yuanyuan Zhang (2019). Pembelajaran jaringan platform pendidikan online dapat membuat integrasi pendidikan dan teknologi informasi yang efektif, dan memiliki dampak besar pada gaya dan mode belajar siswa. Karakteristik belajar siswa dapat direpresentasikan sebagai indikator data yang dapat dioperasikan dan dimengerti melalui data perilaku mereka. Dengan merekam dan menganalisis data perilaku belajar siswa, kita dapat memahami status belajar siswa yang berbeda dan meningkatkan efek belajar mereka. Makalah ini pertama-tama memperoleh indikator pengamatan dengan signifikansi khas sebagai kumpulan data asli, kemudian melalui analisis korelasi untuk mengekstrak enam indikator yang secara objektif dapat mencerminkan karakteristik pembelajaran. Dalam tulisan ini, analisis

kluster indikator-indikator tersebut dilakukan berdasarkan kinerja siswa untuk menggambarkan bahwa pembelajaran online siswa memiliki karakteristik kelompok. Makalah ini berfokus pada karakteristik perilaku belajar dari kelompok yang berbeda, dan memberikan saran yang sesuai dari perspektif manajemen pengajaran [2].

Youguo Li, Haiyan Wu (2012). Menurut analisis akademis dan hasil percobaan di atas, K-Means yang ditingkatkan tidak hanya menjaga efisiensi tinggi dari K-Means standar tetapi juga meningkatkan kecepatan konvergensi secara efektif dengan meningkatkan cara pemilihan titik fokus cluster awal. K-Means yang ditingkatkan jelas lebih baik daripada K-Means standar dalam presisi dan stabilitas cluster. Terutama, keuntungannya akan lebih terlihat ketika sampai pada masalah cluster yang memiliki skala besar dan data terdistribusi secara acak [3].

Kevin Fong-Rey Liu, Jia-Shen Chen (2011). Sistem pendukung keputusan integrasi DM dan BBN dirancang untuk memprediksi hasil belajar siswa. Ini berisi empat langkah: teori fuzzy untuk mengidentifikasi faktor-faktor pada hasil belajar; penambahan data untuk membangun diagram pengaruh; pembelajaran mesin untuk membuat tabel probabilitas; dan model untuk memprediksi nilai ujian di awal kursus dan dengan demikian membantu siswa meningkatkan nilai mereka sesuai dengan kelemahan mereka[4]

Soumi Ghosh, Sanjay Kumar Dubey (2013). Di arena perangkat lunak, teknologi data mining telah dianggap sebagai sarana yang berguna untuk mengidentifikasi pola dan tren volume data yang besar. Pendekatan ini pada dasarnya digunakan untuk mengekstrak pola yang tidak diketahui dari kumpulan besar data untuk bisnis serta aplikasi waktu nyata. Ini adalah disiplin kecerdasan komputasi yang telah muncul sebagai alat yang

berharga untuk analisis data, penemuan pengetahuan baru dan pengambilan keputusan otonom. Data mentah dan tidak berlabel dari kumpulan data dalam jumlah besar dapat diklasifikasikan pada awalnya dengan cara yang tidak diawasi dengan menggunakan analisis kluster yaitu mengelompokkan penugasan sekumpulan pengamatan ke dalam kluster sehingga pengamatan dalam kluster yang sama dapat dianggap serupa. [5]

B. Metode

K-Means Algorithm

Algoritma k-means adalah algoritma untuk mengelompokkan n objek berdasarkan atribut ke dalam k partisi, $k < n$. Hal ini mirip dengan algoritma ekspektasi-maksimalisasi untuk campuran Gauss karena keduanya berusaha menemukan pusat cluster alami dalam data. Diasumsikan bahwa atribut objek membentuk ruang vektor[6]. Tujuannya mencoba untuk mencapai adalah untuk meminimalkan total varians intra-cluster, atau, fungsi kesalahan kuadrat:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

dimana terdapat k cluster S_i , $i = 1, 2, \dots, k$, dan μ_i adalah centroid atau titik rata-rata dari semua titik $x_j \in S_i$. Langkah-langkah algoritmanya adalah:

1. Inisialisasi k pusat cluster menjadi titik benih (Pusat-pusat ini dapat diproduksi secara acak atau menggunakan cara lain untuk menghasilkan).

2. Untuk setiap sampel, cari pusat cluster terdekat, letakkan sampel di cluster ini dan hitung kembali pusat cluster yang diubah (Ulangi n kali).
3. Periksa kembali semua sampel dan masukkan masing-masing ke dalam cluster yang diidentifikasi dengan pusat terdekat (jangan menghitung ulang setiap pusat cluster). Jika anggota setiap cluster belum diubah, berhenti. Jika diubah, lanjutkan ke langkah 2.

Clustering

Clustering adalah salah satu alat untuk menganalisis data, yang memecahkan permasalahan pengelompokan data. Objek dari klustering adalah kasus pendistribusian (orang, objek, peristiwa, angka, atau data lainnya) ke dalam kelompok, sedemikian rupa hingga mendapatkan derajat tingkat keterhubungan antar anggota cluster yang berbeda. Dengan cara ini, masing-masing cluster menguraikan, dalam kaitan dengan kumpulan atau koleksi data, class dimana milik anggotanya. Cluster disebut juga data atau item yang dikelompokkan berdasarkan pilihan konsumen ataupun hubungan logis.

Analisis Kluster ialah metode yang dipakai untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya, Jadi menurut Gorunescu (2011), secara umum dapat dikatakan bahwa :

- a) Data dalam satu cluster memiliki tingkat kesamaan yang tinggi.
- b) Data dalam kluster memiliki tingkat kesamaan yang rendah.

C. Hasil dan pembahasan

Tujuan yang akan dilakukan pada penelitian ini adalah bagaimana melakukan pengelompokan hasil belajar siswa di SMKN 1 Ngasem

Kediri menggunakan metode K-Means. Penelitian dibuat menggunakan data primer, dengan respondennya adalah siswa yang kelas X yang mempunyai nilai produktif yang ada di data sekolah menengah kejuruan negeri 1 ngasem Kediri. Parameter yang digunakan untuk melakukan pengelompokan hasil belajar siswa terhadap pendapatan orangtua siswa berjumlah 2 yaitu Nilai Produktif dan Pendapatan Orang tua siswa

Data yang diolah dalam penelitian ini merupakan sampel yang diambil dari data rapor peserta didik SMKN 1 Ngasem Kediri Kelas X MM tahun ajaran 2019/2020. Dataset siswa terdiri dari atribut data induk dan data rapor. Atribut yang dijadikan variable dalam algoritma ini di sederhanakan menjadi atribut yang terdiri dari Nama, Nilai Teori, Nilai Praktikum, Nilai Produktif dan Pendapatan Orang Tua. Data sampel yang akan diuji cobakan terdiri dari 22 Sampel peserta didik. Selanjutnya dicoba mengelompokkan data diatas menjadi 3 kelompok. Dengan menggunakan algoritma K-means.

Tabel 1. Data Siswa

| No | Name | Practice Score | Theory Score | Productive Score | Income\$ |
|----|---------|----------------|--------------|------------------|----------|
| 1. | Rob | 75 | 76 | 89 | 70 |
| 2. | Michael | 90 | 87 | 78 | 90 |
| 3. | Mohan | 84 | 80 | 85 | 61 |
| 4. | Ismail | 81 | 75 | 85 | 60 |
| 5. | Kory | 76 | 77 | 75 | 150 |
| 6. | ... | ... | ... | ... | ... |
| 7. | ... | ... | ... | ... | ... |
| 20 | Alia | 82 | 75 | 87 | 80 |
| 21 | Sidiq | 81 | 79 | 79 | 82 |
| 22 | Abdul | 86 | 86 | 81 | 58 |

Data Cleaning

Data cleaning ini yang dilakukan adalah membuang atribut yang tidak relevan atau tidak konsisten. Atribut yang dibuang dari dataset yang sudah ada yaitu Practice Score dan Theory Score dan Atribut yang digunakan yaitu Productive Score dan Income dataset yang telah di cleaning dapat dilihat pada tabel 2.

Tabel 2. Data Cleaning

| No | Name | Productive Score | Income\$ |
|----|---------|------------------|----------|
| 1. | Rob | 89 | 70000 |
| 2. | Michael | 78 | 90000 |
| 3. | Mohan | 85 | 61000 |
| 7. | ... | ... | ... |
| 20 | Alia | 87 | 80000 |
| 21 | Sidiq | 79 | 82000 |
| 22 | Abdul | 81 | 58000 |

Data Process Training

Jumlah pusat cluster secara acak (random). Pada percobaan ini ditentukan 3 data secara acak sebagai titik pusat awal untuk perhitungan jarak dari seluruh kelompok cluster yang akan dibentuk.

Jumlah cluster = 3 (sedikit, sedang, tinggi)

Jumlah data = 22

Koordinat cluster pusat yang terbentuk terlihat pada gambar 1. berikut :

```
array([[8.11818182e+01, 5.61363636e+04],
       [8.21428571e+01, 1.50000000e+05],
       [8.32500000e+01, 8.05000000e+04]])
```

results of data normalization

hasil normalisasi variable Productive Score dan Income bisa dilihat pada tabel 3

Tabel 3. Data Normalization

| No | Name | Productive Score | Income\$ |
|----|---------|------------------|----------|
| 1. | Rob | 0,9333 | 0,2111 |
| 2. | Michael | 0,2000 | 0,3846 |
| 3. | Mohan | 0,6666 | 0,1367 |
| 7. | ... | ... | ... |
| 20 | Alia | 0,8000 | 0,2991 |
| 21 | Sidiq | 0,2666 | 0,3613 |
| 22 | Abdul | 0,4000 | 0,1111 |

Setelah melakukan normalisasi angka pada masing-masing variable, berikutnya adalah menentukan jumlah kelompok atau klaster. Sebanyak 22 data siswa akan dikelompokkan menjadi 3 klaster yaitu : Baik, Sedang dan Kurang.

Setelah menentukan jumlah klaster, berikutnya adalah menentukan nilai initial cluster center (centroid) untuk masing-masing klaster pada setiap variabelnya.

Secara sederhana, algoritma K-Means dimulai dari tahap berikut :

1. Memilih K buah titik centroid
2. Menghitung jarak data dengan centroid
3. Update nilai titik centroid

Perhitungan manual akan memakan banyak waktu. Disinilah peran Artificial Intelligent supaya pengolahan data bisa lebih cepat.

Ada banyak bahasa pemrograman yang bisa digunakan untuk penerapan algoritma dalam data mining. Dalam makalah ini peneliti menggunakan bahasa pemrograman python. Dengan fungsi-fungsi yang ada pada library python, akan memudahkan kita dalam penerapan algoritma K-means mulai dari normalisasi data, menentukan jumlah cluster, menentukan nilai centroid per cluster sampai pada pengelompokan data ke dalam cluster yang sudah ditentukan. Penentuan Cluster ada pada tabel 4.

Tabel 4. Data Cluster

| No | Name | Productive Score | Income\$ | Celuster |
|----|---------|------------------|----------|----------|
| 1. | Rob | 0,9333 | 0,2111 | 0 |
| 2. | Michael | 0,2000 | 0,3846 | 2 |
| 3. | Mohan | 0,6666 | 0,1367 | 0 |
| 7. | ... | ... | ... | ... |
| 20 | Alia | 0,8000 | 0,2991 | 0 |
| 21 | Sidiq | 0,2666 | 0,3613 | 2 |
| 22 | Abdul | 0,4000 | 0,1111 | 2 |

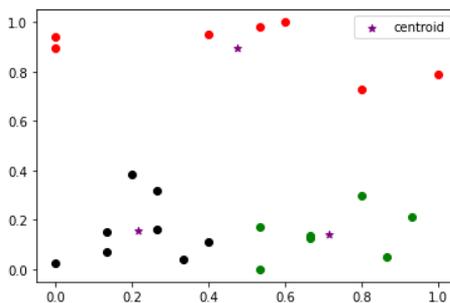
| | StudentName | Practice Score | theoryScore | productivescore | Income(\$) | cluster |
|---|-------------|----------------|-------------|-----------------|------------|---------|
| 0 | Rob | 75 | 76 | 0.933333 | 0.213675 | 0 |
| 1 | Michael | 90 | 87 | 0.200000 | 0.384615 | 2 |
| 2 | Mohan | 84 | 80 | 0.666667 | 0.136752 | 0 |
| 3 | Ismail | 81 | 75 | 0.666667 | 0.128205 | 0 |
| 4 | Kory | 76 | 77 | 0.000000 | 0.897436 | 1 |

Gambar 1. Hasil Cluster Menggunakan Python

Pada gambar 1 terlihat inisial cluster yang ditampilkan dimulai dari angka 0, 1 dan 2. Jadi angka 0 adalah inisial cluster 1, angka 1 adalah inisial cluster 2, angka 2 adalah inisial cluster 3. Sesuai dengan perhitungan

manual yang sudah kita lakukan, data siswa 1 masuk kedalam cluster 0. Hasil dari clustering 22 data siswa kedalam 3 cluster bisa kita lihat pada gambar 2.

- Cluster 1 : Baik (Red)
- Cluster 2 : Sedang (Green)
- Cluster 3 : Kurang (Black)



Gambar 2. Scatter Hasil Clustering 22 Siswa

D. Penutup

Simpulan dan Saran

Hasil penelitian clusterisasi yang berjudul “Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Pendapatan Orang Tua Terhadap Hasil Belajar Pada Mata Pelajaran Produktif” pada data Hasil belajar siswa dapat ditarik kesimpulan bahwa kelompok siswa yang termasuk mendapatkan hasil kurang rata rata orang tua mendapatkan penghasilan pas-pasan ,dan siswa yang mendapatkan hasil baik orang tua siswa adalah golongan dari orang mampu karena untuk menunjang nilai prdofuktif diperlukan alat-alat praktikum yang memadai dan siswa dapat menguasai terlebih kamera dan alat video recorder , sedangkan siswa yang masuk kedalam kelompok sedang orang tua berpenghasilan standar yaitu golongan menengah.

Dari penelitian ini tentu masih banyak terdapat kekurangan sehingga analisa cluster diatas perlu dilakukan lagi penelitian lanjutan agar clusterisasi data obat dapat dilakukan secara lebih valid dengan menetapkan nilai centroid terbaik.

Daftar Pustaka

- Ghosh, Soumi dan Dubey, Sanjay Kumar. 2013. Comparative Analysis of K-Means and Fuzzy C- Means Algorithm. India
- Jian Wang, Yuanyuan Zhang. Clustering Study of Student Groups Based on Analysis of Online Learning Behavior. ICMET 2019, June 28–30, 2019, Nanjing, China.
- Kevin Fong-Rei Liu, Jia-Shen Chen.” Prediction and assessment of student learning outcomes in calculus”. IEEE 2011
- R. Kannan, S. Vempala, and Adrian Vetta, “On Clusterings: Good, Bad, and Spectral”, Proc. of the 41st Foundations of Computer Science, Redondo Beach, 2000.
- Youguo Li, Haiyan Wu 2012. A Clustering Method Based on K-Means Algorithm.
- Yufei Ding, Yue Zhao, Xipeng Shen, Madanlal Musuvathi, and Todd Mytkowicz. 2015. Yinyang k-means: A drop-in replacement of the classic k-means with consistent speedup. In Proceedings of the 32nd International Conference on Machine Learning (ICML-15). 579–587.

