

PREDIKSI PENDAFTARAN PESERTA DIDIK BARU DENGAN METODE POLYNOMIAL REGRESSION, DAN K-MEDOIDS

Noviana¹, Justi Fartesa², Chairani Fauzi³, Sriyanto⁴
Informatics & Business Institute Darmajaya, Bandar Lampung
Jl. 2.A. Pagar Alam No. 93, Bandar Lampung - Indonesia 35142
E-mail : 1Novianagita1989@gmail.com, 2Justifartesa@radenintan.ac.id,
3chairani@darmajaya.ac.id, 4sriyanto@darmajaya.ac.id

ABSTRAK

Salah satu sekolah yang berupaya memprediksi penerimaan peserta didik baru berdasarkan data dari tahun sebelumnya, namun hasilnya tidak tepat. Turun naiknya jumlah penerimaan peserta didik baru ini merupakan masalah bagi SMK Negeri 2 Kotabumi dalam mempersiapkan sarana rombel, seragam, buku penunjang kegiatan pembelajaran serta menentukan langkah- langkah dan kebijakan terkait promosi sekolah dan target penerimaan peserta didik baru tahun-tahun selanjutnya. Prediksi pendaftaran peserta didik baru menggunakan metode polynomial regression, dan K-Medoids, Dalam pengolahan data prediksi pendaftaran peserta didik. Hasil yang didapat nilai Y yang sesuai dengan hasil implementasi menggunakan python. Contohnya pada tahun 2018 didapat nilai $Y = 0,0034x^6 - 0,6194x^5 + 46,754x^4 - 1864,6x^3 + 41412x^2 - 485358x + 2E+06 = 1744,01$ dengan $R = 0,8779$ yang disertai dengan yang sama untuk setiap tahun, sedangkan untuk metode K-Medoids didapat pada tahun 2018 clustering 0 didapat 73 calon siswa dengankategori tidak lulus dan 19 dengan kategori lulus, sedangkan untuk tahun 2019 sampai 2022 jumlah cluster 0 bernilai 0 dan cluster 1 bernilai 92 yangberarti peserta lulus semua.

Kata Kunci : polynomial regression, K-Medoids, Prediksi pendaftaran

ABSTRACT

One school attempted to predict the acceptance of new students based on data from the previous year, but the results were inaccurate. The fluctuation in the number of new student admissions is a problem for SMK Negeri 2 Kotabumi in preparing class facilities, uniforms, books to support learning activities and determining steps and policies related to school promotion and targets for new student admissions in the following years. Predicting new student enrollment using the polynomial regression method, and K-Medoids, in processing student enrollment prediction data. The results obtained are Y values that are in accordance with the implementation results using python. For example, in 2018 the value $Y = 0.0034x^6 - 0.6194x^5 + 46.754x^4 - 1864.6x^3 + 41412x^2 - 485358x + 2E+06 = 1744.01$ with $R = 0.8779$ accompanied by the same for each year, whereas for the K- Medoids method obtained in 2018 clustering 0 obtained 73 prospective students in the non-passing category and 19 in the pass category, while for 2019 to 2022 the number of cluster 0 is worth 0 and cluster 1 is worth 92 which means that all participants have passed.

Keywords : polynomial regression, K-Medoids, Enrollment prediction

1. PENDAHULUAN

Penerimaan peserta didik baru merupakan agenda sekolah atau lembaga pendidikan formal,

hal ini dilaksanakan baik sekolah negeri maupun swasta di waktu yang sama. Proses ini merupakan salah satu kegiatan awal dalam proses kegiatan

belajar mengajar di sekolah. Penerimaan peserta didik baru dilaksanakan di setiap jenjang sekolah, di mulai dari tingkat sekolah dasar, SMP, SMA dan SMK. Dalam setiap tahunnya penerimaan peserta didik baru dapat mengalami kenaikan dapat pula mengalami penurunan. Hal ini dapat mempengaruhi berbagai aspek dalam kegiatan yang telah direncanakan sekolah. Baik dalam persiapan sarana prasarana ataupun anggaran yang sudah diagendakan di tahun ajaran baru, sehingga diperlukan adanya prediksi untuk mengetahui jumlah penerimaan peserta didik baru, agar semua keputusan dalam manajemen perencanaan ke depan dapat berjalan dengan baik. Seni dan ilmu meramalkan peristiwa masa depan dikenal sebagai prediksi[1]. Dimana hasil prediksi ini dapat menjadi panduan untuk

perencanaan kedepannya [2]. Untuk mengumpulkan informasi yang dapat digunakan untuk membuat prediksi, Data Mining dapat digunakan untuk mengekstraksi informasi dari kumpulan data yang sangat besar [3]. Menemukan pola dalam data yang disimpan dalam database di suatu instansi menjadi lebih mudah dengan data mining [4]. Karena keefektifannya, pengelompokan aliran data telah menarik perhatian dalam penelitian. Mengingat bahwa hanya ada satu pindaian, menerapkan teknik pengelompokan acak ke aliran data merupakan tantangan. Pengelompokan aliran dapat dilakukan untuk seluruh aliran atau menggunakan jendela geser di mana hanya item aliran terbaru yang diperhitungkan[5].

SMK Negeri 2 Kotabumi salah satu sekolah yang berupaya memprediksi penerimaan peserta didik baru berdasarkan data dari tahun sebelumnya, namun hasilnya tidak tepat. Turun naiknya jumlah penerimaan peserta didik baru ini merupakan masalah bagi SMK Negeri 2 Kotabumi dalam mempersiapkan sarana rombel, seragam, buku penunjang kegiatan pembelajaran serta menentukan langkah-langkah dan kebijakan terkait promosi sekolah dan target penerimaan peserta didik baru tahun-tahun selanjutnya

2. PENELITIAN YANG TERKAIT

D. A. Ferryan, Dkk. dengan judul “Peramalan Harga Minyak Mentah di Indonesia dengan metode Regresi Polinomial”, menyatakan bahwa peramalan ini memiliki tingkat keakuratan yang baik dan peramalan harga minyak mentah di Indonesia mengalami penurunan dari bulan sebelumnya dimulai dengan harga minyak mentah pada bulan desember 2021 dengan harga US\$ 68,24 /bbl sampai dengan Mei 2022 dengan harga US\$ 48,78/bbl [6]. Sedangkan

Menurut Dyang Falila Pramesti dkk, dengan judul “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)” menyatakan bahwa penentuan potensi diperoleh hasil dimana potensi tinggi terhadap kebakaran hutan/lahan ditunjukkan pada cluster 1 dengan rata-rata brightness diatas batas ambang yaitu 334.470K dan rata-rata confidence 87.08% dengan indikasi api termasuk dalam katagori high confidence fire atau memiliki potensi tinggi terhadap terjadinya kebakaran yang sebenarnya sehingga dibutuhkan penanggulangan segera. Potensi sedang ditunjukkan pada cluster 2 dengan rata-rata brightness dibawah batas ambang yaitu 318.80K dan rata-rata confidence 58.73% dengan indikasi api termasuk dalam katagori nominal confidence fire dimana dibutuhkan kewaspadaan terjadinya kebakaran hutan/lahan[7].

3. METODE PENELITIAN

A. Data Mining

Analisis penemuan data yang dikenal sebagai data berupaya menemukan hubungan yang tidak terduga dan memberi pemilik data ringkasan data yang dapat dipahami dan bermanfaat[8], [9]. Tujuannya adalah untuk menemukan pola yang sebelumnya tidak diketahui. Dalam penambangan data, tumpukan data masa lalu ini dianggap sebagai tambang yang dapat diolah untuk menghasilkan pengetahuan yang berharga[10]–[13].

B. K- Medoids

Ketika pengelompokan dilakukan menggunakan pengelompokan berbasis medoid, metafora Cluster dihitung sebagai jarak rata-rata antara semua item dalam koleksi. Pengelompokan berbasis medoid menganggap medoid yang mewakili satu instance dari kumpulan data sebagai pusat dari setiap Cluster[14]. Algoritma PAM (Partitioning Around Medoids), sering disebut sebagai algoritma KMedoids, adalah metode yang direpresentasikan oleh cluster, khususnya medoid [15].

C. Polynomial Regression

Analisis regresi melibatkan identifikasi hubungan antara variabel dependen dan satu atau lebih variabel independen. Ini adalah salah satu alat statistik terpenting yang banyak digunakan di hampir semua ilmu. Model hubungan dihipotesiskan, dan perkiraan nilai parameter digunakan untuk mengembangkan persamaan regresi yang diperkirakan [16], [17]. Regresi polinomial didasarkan pada premis bahwa ukuran orang dan lingkungan mewakili konstruksi yang berbeda dan asumsi yang tertanam dalam skor perbedaan dan indeks kesamaan profil mewakili hipotesis yang harus diuji secara empiris [18].

4. HASIL DAN PEMBAHASAN

4.1 K-Medoids

Penentuan cluster dibagi menjadi 2 kategori yaitu seperti table dibawah ini:

Tabel I. Penentuan Criteria

Kriteria Cluster	
Cluster 0 - Tidak Lulus	
Cluster 1 - Lulus	

Dari data Clustering pada table didapat hasil Iterasi 1 sebagai berikut:

No	Nama	C0	C1	Kedekatan	Cluster
1	Salsabila Mardiyah	0,87	1,604649494	0,87	0

2	Aisah	0	1,529705854	0	0
3	Anisak Tuljanah	1,56	1,958979326	1,56	0
4	Dani Revaldi	1,017889974	0,700071425	0,700071425	1
5	Celly Inanda	1,122541759	0,542309875	0,542309875	1
6	Pandu Setiawan	2,640075756	1,204159458	1,204159458	1
7	Aldi Pirando	3,146823796	1,634778272	1,634778272	1
...
90	Wulan Pratama Sari	1,770197729	0,64	0,64	1
91	Denisa Fitri Amanda	1,988064385	1,483374531	1,483374531	1
92	Beta Agus Setiyana	3,097240707	1,571909667	1,571909667	1
Jumlah Kedekatan				734,0188709	

Didapat utk data tahun 2018 sampai batas iterasi 3 sehingga didapat hasil seperti table dibawah ini:

No	Nama	C0	C1	Kedekatan	Cluster
1	Salsabila Mardiyah	3,141562498	3,491227572	3,141562498	0
2	Aisah	2,473876094	2,719049459	2,473876094	0
3	Anisak Tuljanah	3,730876429	4,131385961	3,730876429	0
4	Dani Revaldi	1,716402904	2,171301444	1,716402904	0
5	Celly Inanda	2,361727954	2,861459411	2,361727954	0
6	Pandu Setiawan	2,887521936	3,559594072	2,887521936	0
7	Aldi Pirando	2,936048864	3,634175279	2,936048864	0
...

90	Wulan Pratama Sari	2,6743199	3,262071422	2,6743199	0
91	Denisa Fitri Amanda	0,568484767	1,155677273	0,568484767	0
92	Beta Agus Setiyana	2,7733833	3,473049087	2,7733833	0
Jumlah Kedekatan				1089,855901	

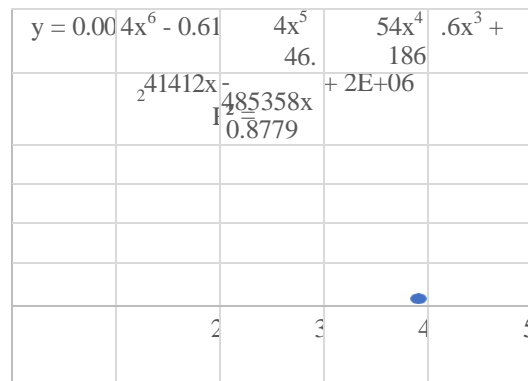
3,85	8,40	14,82	57,07	21,971	8,40	32,034	124,51
4,00	9,00	16,00	64,00	25,000	9,00	36,000	144,00
3,80	8,00	14,44	54,87	20,851	8,00	30,40	115,52
4,64	8,40	21,53	99,90	46,352	8,40	38,098	180,85
4,71	7,00	22,18	104,92	49,13	7,00	32,097	155,29
43,00	72,80	187,57	831,81	37,360	31,72	13,258	13,94

Dari keseluruhan perhitungan menggunakan metode K- Medoids didapat seperti table dibawah ini:

Tahun	Jumlah Iterasi	Jumlah Cluster 0	Jumlah Cluster 1
2018	3	73	19
2019	2	0	92
2020	2	0	92
2021	3	0	92
2022	3	0	92

Dengan penyelesaian matriks maka didapat nilai $y = 72,80$

$$+ 312,27 + 1358,94 = 1744,01$$



4.2 Polynomial Regression

Tabel 28 Tabel Penyelesaian

Xn	Yn	Xn ²	Xn ³	Xn ⁴	Xn ⁵	Xn ⁶	Xn ⁷
4,08	6,00	16,64	67,72	27,68	6,00	24,00	99,88
4,95	6,00	24,50	90,09	60,00	6,00	29,70	147,02
3,39	6,00	11,48	38,60	13,07	6,00	20,00	68,95
5,14	7,00	26,42	135,00	69,00	7,00	35,00	184,94
4,44	7,00	19,72	87,37	38,00	7,00	31,00	138,00

Proses perhitungan kesesuaian dalam pendaftaran siswa baru di SMKN 2 Kotabumi menggunakan polynomial regression yang didapat dari prediksi pendaftaran siswa baru pada tahun 2018-2022 adalah sebagai berikut:

1. Prediksi pendaftaran peserta didik baru
2. Total prediksi pendaftaran peserta didik baru

- Prediksi berdasarkan nilai tes pada tahun 2018 = 1.744,01

- Prediksi berdasarkan nilai tes pada tahun 2019 = 2.098,43

- Prediksi berdasarkan nilai tes pada tahun 2020 = 4468,72

- Prediksi berdasarkan nilai tes pada tahun 2021 = 4801,27

Prediksi berdasarkan nilai tes pada tahun 2022 = 6.018,13

5. KESIMPULAN

Berdasarkan perhitungan menggunakan Polinomial Regrasion didapat nilai Y yang sesuai berdasarkan alur grafik yang didapat dengan hasil $R = 0,8779$ untuk tahun 2018, begitu juga untuk tahun selanjutnya sampai tahun 2022 dengan hasil yg sama berdasarkan pergerakan grafiknya.

Sedangkan untuk metode K-Medoids didapat pada tahun 2018 clustering 0 didapat 73 calon siswa dengankategori tidak lulus dan 19 dengan kategori lulus, sedangkan untuk tahun 2019 sampai 2022 jumlah cluster 0 bernilai 0 dan cluter 1 bernilai 92 yangberarti peserta lulus semua. Dari kedua metode dapat disimpulkan dengan hasil yang sama seimbang untuk data tahun 2019- 2022, sedangkan utk 2018 berbeda antara kedua metode ini.

PUSTAKA

[1] “Prediksi Pro duksi Daging Sapi Nasional dengan Meto de Regresi Linier dan Regresi Polinomial,” *Jurnal Ilmiah Komputasi*, vol. 20, no. 2, Jun. 2021, doi: 10.32409/jikstik.20.2.2722.

[2] D. Suwardiyanto, M. Nur Shodiq, D. Hidayat Kusuma, and T. Oktalita Sari, “Sistem Prediksi Kebutuhan Obat di Puskesmas Menggunakan Metode Least Square,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 1,

pp. 75–80, Jan. 2019, doi: 10.30591/jpit.v4i1.1085.

[3] H. Thamrin, “Analyzing and Forecasting Admission data using Time Series Model,” *Jurnal Online Informatika*, vol. 5, no. 1, pp. 35–44, 2020, doi: 10.15575/join.v5i1.546.

[4] A. Afandi, N. Noviana, and D. Nurdianah, “Naive Bayes Method and C4.5 in Classification of Birth Data,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 16, no. 4, p. 435, Oct. 2022, doi: 10.22146/ijccs.78198.

[5] E. Alothali, H. Alashwal, and S. Harous, “Data stream mining techniques: A review,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 2, pp. 728–737, Apr. 2019, doi: 10.12928/TELKOMNIKA.v17i2.11752.

[6] D. A. Ferryan, P. K. Intan, and M. Hafiyusholeh, “Peramalan Harga Minyak Mentah di Indonesia dengan Metode Regresi Polinomial,” *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, vol. 19, no. 1, pp. 13–18, Jun. 2022, doi: 10.22487/2540766x.2022.v19.i1.15779.

[7] P. Kebakaran Hutan, L. Berdasarkan Persebaran Titik Panas Dyang Falila Pramesti, M. Tanzil Furqon, and

C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data,” 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>

- [8] E. Vivian Astuti, A. Afandi, and D. Marisa Efendi, "Classification and Clustering of Internet Quota Sales Data Using C4.5 Algorithm and K-Means ARTICLE INFORMATION ABSTRACT," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 2, pp. 268–283, 2023, doi: 10.26555/jiteki.v9i2.25970.
- [9] K. Natarajan, J. Li, and A. Koronios, "Data mining techniques for data cleaning," *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management, WCEAM 2009*, no. September, pp. 796–804, 2009, doi: 10.1007/978-0-85729-320-6_91.
- [10] Xiaoling Shu, *Knowledge Discovery in the Social Sciences: A Data Mining Approach*. Oakland, California: University of California Press, 2020.
- [11] R. Muliono and Z. Sembiring, "DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT," vol. 4, no. 2, pp. 272–279, 2019.
- [12] R. Hasan, S. Palaniappan, S. Mahmood, K. U. Sarker, and A. Abbas, "Modelling and predicting student's academic performance using classification data mining techniques," *Int J Bus Inf Syst*, vol. 34, no. 3, pp. 403–422, 2020, doi: 10.1504/IJBIS.2020.108649.
- [13] T. R. Stella Mary and S. Sebastian, "Predicting heart ailment in patients with varying number of features using data mining techniques," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 2675–2681, 2019, doi:10.11591/ijece.v9i4.pp2675-2681.
- [14] A. M. Jabbar and K. R. Ku-Mahamud, "Grey wolf optimization algorithm for hierarchical document clustering," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, pp. 1744–1758, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1744-1758.
- [15] D. Transaksi Bongkar Muat di Provinsi Riau, I. Kamila, U. Khairunnisa, P. Studi Sistem Informasi, and F. Sains dan Teknologi UIN Sultan Syarif Kasim Riau, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.
- [16] E. Ostertagová, "Modelling using polynomial regression," in *Procedia Engineering*, Elsevier Ltd, 2012, pp. 500–506. doi: 10.1016/j.proeng.2012.09.545.
- [17] V. Fibriyani and N. Chamidah, "Prediction of Inflation Rate in Indonesia Using Local Polynomial Estimator for Time Series Data," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1776/1/012065.
- [18] J. R. Edwards, "Polynomial regression and response surface methodology 1 Polynomial Regression and Response Surface Methodology."