

PERBANDINGAN ALGORITMA C4.5 DAN REGRESI LINEAR DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN UANG KULIAH

¹Ayu Ratna Juwita, ²Adi Rizky Pratama, ³Tohirin Al Mudzakir, ⁴Cici Emilia Sukmawati
Program Studi Teknik Informatika, Universitas Buana Perjuangan
Email: 1ayurj@ubpkarawang.ac.id, 2adi.rizky@ubpkarawang.ac.id,
3tohirin@ubpkarawang.ac.id, 4cici.emilia@ubpkarawang.ac.id

ABSTRAK

Universitas merupakan institusi perguruan tinggi yang memberikan suatu gelar tingkat sarjana dalam berbagai bidang akademik pada dunia pendidikan. Salah satu kewajiban untuk peserta didik untuk mendapatkan pendidikan adalah melakukan pembayaran Sumbangan Pembangunan Pendidikan (SPP). Pembayaran uang kuliah merupakan data deret waktu yang dapat dianalisis sebagai bahan penelitian karena perubahan datanya tidak terlalu cepat, dimana data tersebut akan mengalami trend tertentu baik itu naik maupun turun. Berdasarkan penjelasan tersebut tujuan penelitian ini yaitu untuk mengetahui nilai prediksi keterlambatan pembayaran mahasiswa. Metode penelitian yang digunakan yaitu algoritma regresi linier dan algoritma C.45 dengan perhitungan tingkat error menggunakan metode *Mean Square Error* (MSE) dan *Mean Absolute Error* (MAE). Dari hasil perhitungan pengujian yang dilakukan maka didapatkan tingkat error yang lebih kecil dihasilkan dari algoritma regresi linier sedangkan algoritma C.45 memiliki nilai tingkat error yang lebih tinggi dibandingkan algoritma regresi linier. Berdasarkan hasil tersebut didapatkan sebuah kesimpulan bahwa algoritma regresi linier untuk perhitungan prediksi keterlambatan pembayaran mahasiswa dinilai lebih baik dibandingkan dengan algoritma C.45.

Kata Kunci : *Peramalan, Prediksi, Algoritma C.45, Regresi Linear.*

PENDAHULUAN

Universitas merupakan institusi perguruan tinggi yang memberikan suatu gelar akademisi dalam berbagai bidang pendidikan, salah satunya Universitas Buana Perjuangan Karawang didirikan oleh Yayasan Pembina Perguruan Tinggi Pangkal Perjuangan (YPPTPP) pada tahun 2014. UBP Karawang mendapatkan SK pendirian dari Menteri Pendidikan dan Kebudayaan Republik Indonesia melalui SKI Nomor 611/E/O/2014 Tanggal 17 Oktober 2015 tentang Izin Pendirian Universitas Buana Perjuangan Karawang, terdiri dari 6 Fakultas diantaranya Fakultas Ekonomi dan Bisnis, Fakultas Teknik dan Ilmu Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Fakultas Psikologi, Fakultas Hukum dan Fakultas Farmasi.

Seiring pesatnya kemajuan ilmu pengetahuan, kesadaran mengenai peristiwa yang akan

datang semakin bertambah dan kebutuhan pada berbagai peramalan/prediksi semakin meningkat. Misalnya prediksi di bidang ekonomi, pendidikan, perdagangan, industri, lingkungan dan sosial telah menghadirkan berbagai macam hasil yang dapat digunakan oleh beragam pihak untuk mengambil keputusan. Sering dijumpai berbagai masalah yang bersifat musiman disekitar kehidupan kita. Permasalahan tersebut bagi sebagian orang mungkin hanya menjadi angin lalu yang tidak perlu untuk diperhatikan, padahal jika dicermati dan diteliti, pola musiman yang memiliki pola berulang-ulang tersebut dapat memberikan suatu gambaran kondisi masa depan yang dapat dibuat dalam suatu perencanaan dan pengambilan suatu keputusan yang baik berdasarkan peramalan yang telah dilakukan.

Untuk memprediksi pembayaran uang kuliah disetiap tahunnya, dapat digunakan analisis deret waktu (*time series*). pembayaran uang kuliah merupakan data deret waktu yang perubahannya tidak terlalu cepat, jumlah pembayaran uang kuliah setiap tahun akan mengalami *trend* tertentu, apakah jumlahnya naik atau turun. Metode peramalan ini didasarkan atas konsep bahwa hasil observasi saat ini dipengaruhi oleh hasil observasi masa lalu dan hasil observasi yang akan datang dipengaruhi oleh hasil observasi saat ini.

Pendidikan merupakan salah satu kebutuhan penting bagi setiap orang. Salah satu kewajiban untuk mendapatkan pendidikan di sekolah adalah melakukan pembayaran Sumbangan Pembangunan Pendidikan (SPP). Tidak dapat dihindari bahwa SPP merupakan kewajiban siswa untuk membayarnya dan juga merupakan faktor penting untuk mengalokasikan biaya pembangunan, biaya pengajar, karyawan dan lain sebagainya. (Ginting, Kusrini, and Luthfi 2020).

Muhartini et al. 2021 melakukan penelitian menyatakan hasil prediksi jumlah penerimaan mahasiswa baru dari prodi Manajemen menggunakan metode regresi linear hasil pengujian tingkat error sebesar 3,444% atau tingkat akurasi 96,556%. Pada penelitian yang dilakukan oleh (Ginting, Kusrini, and Luthfi 2020) menghasilkan tingkat akurasi sebesar 73% dalam prediksi keterlambatan pembayaran uang sekolah dengan menggunakan algoritma C4.5. Sedangkan penelitian yang telah dilakukan (Muqorobin, Kusrini, and Luthfi 2019) dalam penelitiannya “Optimasi Metode *Naive Bayes* dengan *Feature Selection Gain* Untuk Memprediksi Keterlambatan Pembayaran Uang Sekolah” dapat menghasilkan tingkat akurasi 80%, sedangkan dalam penerapan optimasi dengan menggunakan *information gain* dan *Naive Bayes* menghasilkan tingkat akurasi sebesar 90%.

(Prabowo Budi Utomo, Ema Utami 2019) dalam penelitiannya dengan menambahkan metode *Regresi Linear* pada metode K-NN dapat meningkatkan tingkat prediksi dari algoritma tersebut, hal ini ditunjukkan dengan semakin kecilnya nilai RMSE yang diperoleh. Penentuan

nilai parameter K yang digunakan akan menentukan nilai yang dihasilkan pada proses selanjutnya yaitu sebesar 0.05807 dan nilai *relative error* sebesar 0.39%. Sedangkan penelitian yang dilakukan oleh (Sucipto 2015) Algoritma klasifikasi data mining dengan model algoritma C4.5 dilakukan dengan pengujian-pengujian yang terukur melalui uji AUC, ROC dan T-Test dengan bantuan *rapid miner* yang telah dilakukan pengujian dengan menggunakan parameter biodata nasabah sebanyak 1312 data, yang menghasilkan akurasi sebesar secara keseluruhan nilai hasil validasi yaitu *accuracy* = 91,06%, *precision* = 100,00% dan *recall* = 78,00%. Maka dari itu penelitian ini membahas tentang bagaimana memprediksi/peramalan pembayaran uang kuliah mahasiswa di Universitas Buana Perjuangan Karawang dengan perbandingan model prediksi/peramalan dengan metode C4.5 dan *regresi linier* untuk mengetahui hasil keterlambatan pembayaran dari metode tersebut. Pada proses peramalan dengan menggunakan persamaan C.45 dan *regresi linier* sebagai berikut:

a. Persamaan C.45

Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal.

Berikut merupakan rumus *information gain* yaitu (Widaningsih 2019):

$$Gain(S, A) = Entropy(S) - \sum_{I \in I} \frac{|S_v|}{|S|} entropy(S_v) \quad (1)$$

Dimana :

A : atribut

|S_v| : jumlah sampel untuk nilai v

|S| : jumlah seluruh sampel data

Entropi adalah keberagaman suatu data. Pada persamaan (2) merupakan rumus entropi:

$$Entropy(S) = - \sum p_i \log_2 p_i \quad (2)$$

Dimana : p_i = porsi atau rasio antara jumlah sampel kelas i dengan jumlah semua sampel pada himpunan data.

b. Persamaan *Regresi Linear*

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen; respon; Y) dengan satu atau lebih variabel bebas (independen, prediktor, X).

Model analisis regresi linear sederhana:

$$Y = \alpha + bx \quad (3)$$

Dimana:

Y = variabel dependen

α = konstanta

b = koefisien regresi

Komponen pada Linear Regresi ada tiga yaitu a sebagai intercept, b sebagai *slope* dan x sebagai indeks waktu. Persamaan untuk mendapatkan nilai a dan b adalah:

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (4)$$

$$\alpha = \frac{\sum y - b \sum x}{n}$$

(5)

Langkah-langkah metode yang disusulkan berdasarkan Linear Regresi sebagai berikut:

1. Pembuatan dataset

2. Pembentukan model linier regresi Langkah pembentukan model sebagai berikut:

a. Langkah 1: Hitung X², Y², XY dan total dari masing-masingnya

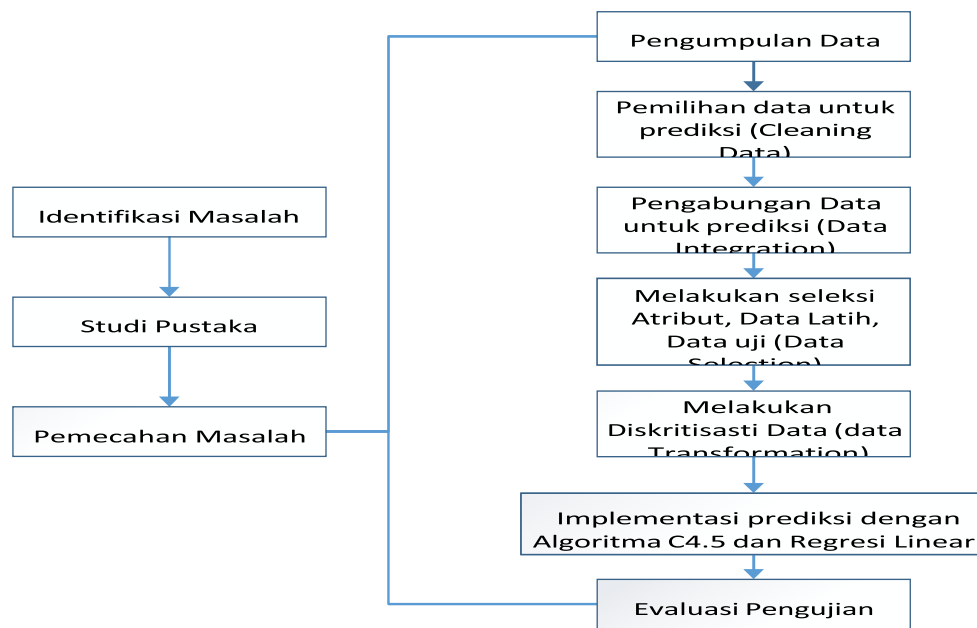
b. Langkah 2: Hitung a dengan menggunakan persamaan (7) dan b menggunakan persamaan (8)

c. Langkah 3: Buat model persamaan linier regresi sederhana

d. Langkah 4: Lakukan prediksi atau peramalan terhadap variabel faktor penyebab atau variabel akibat. (Muhartini et al. 2021)

METODE PENELITIAN

Dalam metode penelitian ini menjelaskan prosedur pemodelan dan mengimplementasikan algoritma C.45 dan *Regresi Linear* untuk memprediksi keterlambatan pembayaran uang kuliah. Sumber dan jenis data yang digunakan dalam pengumpulan data dengan menganalisis data pembayaran mahasiswa di Universitas Buana Perjuangan Karawang. Subjek dalam penelitian ini yaitu membuat sebuah model untuk memprediksi data pembayaran uang kuliah.



Gambar 1. Diagram Penelitian

1. Pemecahan Masalah

Dalam melakukan pemecahan masalah berdasarkan identifikasi dan rumusan masalah yang telah dilakukan untuk memprediksi keterlambatan pembayaran uang kuliah yaitu menggunakan pendekatan komputasi dan memanfaatkan teknik komputasi menggunakan model algoritma C4.5 dan *Regresi Linear*.

a. Pengumpulan Data

Tahap awal yang dilakukan untuk pengumpulan data pada penelitian ini adalah mempersiapkan data pembayaran mahasiswa di Universitas Buana Perjuangan Karawang, data tersebut akan dijadikan dataset sebagai bahan penelitian.

b. Pemilihan Data untuk Predikasi (*Data Cleaning*)

Pada tahap ini dilakukan proses pemilihan data yang akan digunakan dalam penelitian. Proses ini meliputi pembersihan data untuk menghilangkan *attribut*, *noise* dan data yang bersifat inkonsisten,

c. Pengabungan Data (*Data Integration*)

Integrasi data dilakukan untuk menggabungkan data dari beberapa tabel yang berbeda.

d. Seleksi Data (*Data Selection*)

Pada tahap ini data akan diseleksi untuk Melakukan seleksi terdiri atribut, data Latih dan data uji.

e. Transformasi Data (*Data Transformation*)

Transformasi data dilakukan untuk menggabungkan data ke dalam bentuk yang sesuai untuk proses data prediksi.

f. Implementasi prediksi dengan Algoritma C4.5 dan *Regresi Linear*

Pada tahap ini akan dibuatkan sebuah model peramalan menggunakan metode C5.5 dan *regresi linear* untuk memprediksi jumlah keterlambatan pembayaran uang kuliah pada setiap tahun ajarannya.

2. Evaluasi hasil pengujian

Untuk mengetahui apakah metode yang dikembangkan sesuai dengan analisis yang telah dilakukan dan dapat memberikan hasil analisis serta perhitungan yang tepat dalam menentukan hasil prediksi jumlah keterlambatan pembayaran uang kuliah.

Maka dilakukan pengujian sistem untuk mengetahui akurasi dari model yang dikembangkan. Pengujian sistem didasarkan pada jumlah sampel *dataset* yang sudah disiapkan.

HASIL PENELITIAN DAN PEMBAHASAN

1. Pengumpulan Data Pembayaran Kuliah

Data pembayaran kuliah yang digunakan pada penelitian ini sebagai berikut :

Tabel 1 . Data Pembayaran Kuliah

Tahun	Kode Prodi	NIM	MHS	Bayar C1 (%)	Bayar C2 (%)	Bayar C3 (%)	Bayar C4 (%)
2015	001	150001	Mhs001-A	1	1	1	1
2015	001	150002	Mhs001-B	1	1	0	0
2015	001	150003	Mhs001-C	1	0	0	0
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
2015	010	151277	Mhs010-A	1	1	1	0
2020	010	204168	Mhs010-B	1	1	0	0

Sumber : Bagian Keuangan Universitas Buana Perjuangan Karawang

Keterangan Bayar Cicilan

Lunas : 1

Belum Lunas : 2

2. Data Cleaning, Data Integration, Data Selection, Data Transformation

Berikut hasil dari proses *Data Cleaning*, *Data Integration*, *Data Selection*, *Data Transformation* pada data pembayaran kuliah.

Tahun	Jumlah Mahasiswa	Bayar C1 (%)	Bayar C2 (%)	Bayar C3 (%)	Bayar C4 (%)
2015	1277	99.69	93.97	88.57	83.95
2016	2811	93.77	86.91	83.81	80.29
2017	4575	88.87	83.93	80.46	76.81
2018	6131	91.27	87.33	81.91	80.18
2019	6387	94.22	89.43	85.44	82.98
2020	8926	94.06	88.31	85.69	79.42

3. Hasil perhitungan prediksi

a. Prediksi keterlambatan pembayaran menggunakan regresi linear

Berikut hasil dari proses penerapan algoritma *Regresi Linear* menggunakan *python*.

Tahapan Pembayaran	Tahun			Tingkat Error		
	2021	2022	2023	R-Squared	MSE	MAE
C1	7.939 %	7.925 %	7.884 %	-2.09	4.83	2.18
C2	13.533 %	13.521 %	13.484 %	-0.72	1.89	1.33
C3	16.827 %	16.817 %	16.789 %	0.89	0.68	0.75
C4	21.204 %	21.193%	21.161 %	0.53	4.52	2.12

b. Prediksi keterlambatan pembayaran menggunakan C.45

Berikut hasil dari proses penerapan algoritma *C.45* menggunakan *python*.

Tahapan Pembayaran	Tahun			Tingkat Error		
	2021	2022	2023	R-Squared	MSE	MAE
C1	9 %	9 %	9 %	0.00	9.00	3.00
C2	13 %	13 %	13 %	-9.00	2.50	1.50
C3	16 %	16 %	16 %	-9.00	2.50	1.50
C4	20 %	20 %	20 %	-0.25	5.00	2.00

KESIMPULAN DAN IMPLIKASI

Dari hasil penelitian yang telah dilakukan :

1. Algoritma regresi linier dapat digunakan untuk melakukan prediksi pembayaran kuliah mahasiswa, dengan tingkat *error* yang dihasilkan lebih kecil
2. Algoritma C 4.5 dapat diterapkan untuk melakukan perhitungan prediksi keterlambatan pembayaran kuliah mahasiswa namun dengan tingkat *error* yang dihasilkan lebih besar dibandingkan dengan algoritma regresi linier.
3. Berdasarkan penerapan kedua algoritma diatas maka dapat disimpulkan hasil perhitungan prediksi dengan algoritma regresi linier memiliki tingkat error lebih kecil.

Implikasi dari penelitian ini yaitu dapat menggunakan algoritma selain yang digunakan diatas, agar dapat diketahui metode dan algoritma yang lebih tepat untuk digunakan pada penelitian ini.

DAFTAR PUSTAKA

- Ginting, Victor Saputra, Kusrini Kusrini, And Emha Taufiq Luthfi. 2020. "Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python." *Jurnal Teknologi Informasi* 4 (1): 1–6. <https://doi.org/10.36294/Jurti.V4i1.1101>.
- Muhartini, Ajeng Afifah, Oman Sahroni, Septi Dwi Rahmawati, Tanti Febrianti, And Isnaini Mahuda. 2021. "Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana." *Jurnal Bayesian : Jurnal Ilmiah Statistika Dan Ekonometrika* 1 (1): 17–23. <http://bayesian.lppmbinabangsa.id/index.php/home/article/view/2>.
- Muqorobin, Muqorobin, Kusrini Kusrini, And Emha Taufiq Luthfi. 2019. "Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah." *Jurnal Ilmiah Sinus* 17 (1): 1. <https://doi.org/10.30646/Sinus.V17i1.378>.
- Prabowo Budi Utomo, Ema Utami, Suwanto Raharjo. 2019. "Mplementasi Metode K-Nearest Neighbor Dan Regresi Linear Dalam Prediksi Harga Emas." *Jurnal Informasi Interaktif* 4 (3): 155.
- Sucipto, Adi. 2015. "Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Alogaritma Klasifikasi C4.5." *Jurnal Disprotek* 6 (1): 75–87.
- Widaningsih, Sri. 2019. "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm." *Jurnal Tekno Insentif* 13 (1): 16–25. <https://doi.org/10.36787/Jti.V13i1.78>.