

# Prediksi Nilai Akhir Mahasiswa dengan Metode Regresi (Studi Kasus Mata Kuliah Pemrograman Dasar)

Anita Qoiriah<sup>1</sup>, Yuni Yamasari<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Negeri Surabaya

<sup>1</sup>[anitaqoiriah@universitas.ac.id](mailto:anitaqoiriah@universitas.ac.id)

<sup>2</sup>[yuniyamasari@universitas.ac.id](mailto:yuniyamasari@universitas.ac.id)

**Abstrak**— Evaluasi merupakan salah satu cara untuk mengetahui hasil pembelajaran seorang mahasiswa dan membantu mahasiswa dalam mengetahui kinerja pembelajarannya. Untuk membantu mahasiswa dan dosen dalam melacak kemajuan pembelajaran, diperlukan adanya prediksi kinerja mahasiswa sebelum pembelajaran berakhir. Tujuan dari prediksi adalah untuk memperkirakan nilai variabel yang tidak diketahui yang menggambarkan kinerja mahasiswa. Dalam Pendidikan, nilai-nilai yang biasanya diprediksi adalah kinerja, pengetahuan, skor atau tandai. Nilai ini dapat berupa nilai numerik / kontinu atau nilai kategori / diskrit. Analisis regresi menemukan hubungan antara variabel terikat dan satu atau lebih variabel bebas. Penelitian ini dimaksudkan untuk memprediksi nilai akhir mahasiswa peserta matakuliah Pemrograman Dasar menggunakan algoritma regresi. Variabel terikat adalah nilai akhir dan variabel bebas bervariasi dari 1 nilai tugas teori, 2 nilai tugas praktikum dan 1 nilai UTS. Data merupakan hasil pembelajaran sejumlah 105 mahasiswa. Data akan dibagi 2 secara acak. 75% data digunakan sebagai data *training* dan 25% data sebagai data *testing*. Dari 4 model yang dibuat, model 3 mempunyai nilai RMSE yang terbaik yaitu 1.949. Pada model 3 ini menggunakan prediktor nilai tugas, 2 nilai praktikum serta nilai UTS. Sehingga dapat diambil kesimpulan bahwa ketiga jenis nilai tersebut sangat berpengaruh terhadap kinerja mahasiswa pada pembelajaran matakuliah Pemrograman Dasar.

**Kata Kunci**— prediksi, Regresi, RMSE, data *training*, data *testing*.

## I. PENDAHULUAN

Selama beberapa dekade terakhir, pedagogi pendidikan telah banyak berkembang. Dengan munculnya Pembelajaran Sistem Manajemen (LMS), menjadi lebih mudah untuk menyampaikan materi pelajaran kepada mahasiswa. Tapi tetap saja sulit untuk memprediksi kinerja mahasiswa. Memprediksi kinerja mahasiswa dapat membantu mahasiswa dan dosen untuk melacak kemajuan pembelajaran mahasiswa. Melakukan evaluasi merupakan salah satu cara untuk mengetahui hasil pembelajaran seorang mahasiswa. Tujuan dari evaluasi adalah untuk membantu mahasiswa dalam mengetahui kinerja pembelajarannya. Dalam sistem evaluasi beberapa tes dan tugas dilakukan secara berkala. Keberhasilan pembelajaran seorang mahasiswa diharapkan akan terlihat pada hasil tes maupun tugas-tugasnya.

Teknik Data Mining banyak digunakan dibidang pendidikan untuk menemukan pola baru dan / atau tersembunyi data siswa[1][2][3]. Pola tersembunyi ini kemudian dapat digunakan memahami beberapa masalah yang

muncul dalam manajemen pendidikan. Tujuan utama dari *Educational Data Mining* (EDM) ini adalah untuk mempelajari fitur model yang penting untuk memprediksi prestasi akademik mahasiswa. Teknik prediksi Data Mining digunakan untuk mengklasifikasikan faktor yang paling efektif untuk menentukan prestasi akademik mahasiswa.

Penerapan teknik data mining untuk memprediksi kinerja akademis akan memandu dosen dan mahasiswa untuk mencapai prestasi akademik yang lebih baik. Komponen utama data mining menerapkan berbagai teknik dan perhitungan untuk mengekstrak dan mengidentifikasi fitur dari sejumlah besar data. Dari penelitian [4] dengan data mining didapatkan bahwa latihan / proyek lab adalah atribut paling kritis yang sangat mempengaruhi prestasi akademik siswa diikuti dengan kuis, ujian akhir dan ujian tengah semester pada matakuliah Organisasi Komputer. Serta mampu mengidentifikasi siswa yang akan lulus dengan akurasi 89.0%, gagal pada Akurasi 92,60% dan bersyarat pada 74,90%.

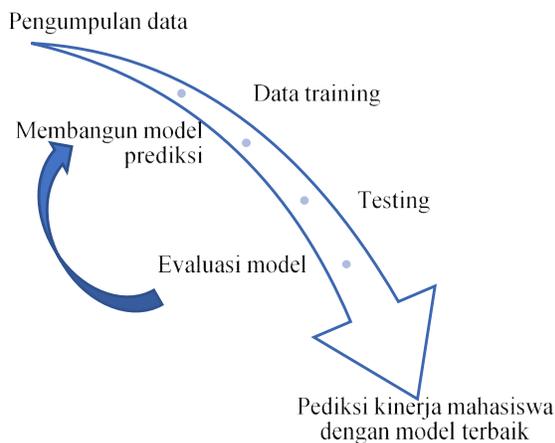
Tujuan dari prediksi adalah untuk memperkirakan nilai variabel yang tidak diketahui yang menggambarkan kinerja mahasiswa. Dalam pendidikan, nilai-nilai yang biasanya diprediksi adalah kinerja, pengetahuan, skor atau tandai. Nilai ini dapat berupa nilai numerik / kontinu atau nilai kategori / diskrit. Analisis regresi menemukan hubungan antara variabel terikat dan satu atau lebih variabel bebas[5]. Penelitian ini dimaksudkan untuk memprediksi nilai akhir mahasiswa peserta matakuliah Pemrograman Dasar menggunakan metode regresi.

## II. METODOLOGI PENELITIAN

Dalam melakukan penelitian ini, maka langkah-langkah yang dilakukan seperti pada gambar 1 yaitu:

- Mengumpulkan data yang digunakan dalam proses prediksi
- Mempersiapkan data *training*
- Membangun model prediksi dengan prediktor tertentu
- Melakukan uji terhadap model dengan data *testing*
- Mengevaluasi hasil uji dengan Root Mean Squared Error (RMSE)

Jika hasil uji belum tidak memuaskan, maka kembali membangun model lain dengan prediktor yang berbeda. Model dengan nilai RMSE terendah merupakan model terbaik yang didapatkan.



Gbr. 1 Langkah-langkah penelitian

### A. Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah data primer. Data primer ini berupa nilai yang didapat dari hasil pembelajaran mahasiswa peserta matakuliah Pemrograman Dasar dalam satu semester di jurusan Teknik Informatika Unesa. Data berupa nilai-nilai tugas baik teori maupun praktik serta nilai ujian tengah semester (UTS).

### B. Data Training dan Data Testing

Untuk memilih variabel relevan yang bisa digunakan untuk memprediksi kinerja mahasiswa secara efektif, maka atribut potensial yang digunakan adalah 1 nilai tugas teori, 2 nilai tugas praktikum dan 1 nilai UTS.

Data didapat dari hasil pembelajaran sejumlah 105 mahasiswa. Dan selanjutnya dilakukan *preprocessing data* agar didapatkan data yang siap untuk digunakan dalam proses prediksi.

Setelah didapatkan data yang siap untuk dimasukkan dalam model regresi, maka data akan dibagi 2 secara acak. 75% data digunakan sebagai data *training* dan 25% data sebagai data *testing*.

### C. Regresi Linier

Regresi linier mengasumsikan bahwa hubungan antara variabel terikat dan variabel bebas dapat dijelaskan secara linier. Targetnya adalah menemukan estimasi nilai yang terbaik untuk koefisien persamaan dengan menggunakan nilai-nilai variabel bebas. Pendekatan standar dalam mencari koefisien adalah menerapkan metode *least-squares*[6].

Regresi linier berganda dapat diartikan sebagai hubungan antara variabel terikat dan dua atau lebih bebas variabel. Asumsi teoritis adalah, untuk setiap perubahan satu unit dalam variabel bebas, selalu ada dan perubahan seragam dalam variabel terikat. Model untuk regresi linier berganda adalah sebagai berikut[7]:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon \quad \dots(1)$$

Dimana Y adalah variabel terikat,  $X_{i1}, X_{i2}, \dots, X_{ik}$  adalah variabel bebas.  $\beta_0$  adalah nilai konstanta dan  $\beta_1, \beta_2 \dots \beta_k$  adalah koefisien regresi.

Variabel terikat (keluaran Y) dari model regresi adalah nilai akhir mahasiswa dalam matakuliah Pemrograman Dasar. Prediktor atau variabel bebas (yaitu, input  $X_1, X_2, X_3$ , dll.) adalah

- $X_1$  : nilai tugas
- $X_2$  : nilai praktikum sebelum UTS
- $X_3$  : nilai praktikum setelah UTS
- $X_4$  : nilai UTS

### D. Evaluasi Model

Root Mean Squared Error (RMSE) adalah metode untuk mengukur kesalahan besaran rata-rata yang dibuat oleh model pada saat memprediksi hasil observasi. Dengan menggunakan RMSE dapat diidentifikasi model mana yang memberikan hasil akurasi tinggi dalam melakukan prediksi [8]. RMSE merupakan akar kuadrat dari rata-rata nilai perbedaan antara nilai yang sebenarnya dan nilai prediksi. Semakin rendah nilai RMSE, semakin baik modelnya dan semakin akurat hasilnya.

Secara matematis, rumusnya ditulis sebagai berikut.

$$RMSE = \left( \frac{\sum (y_i - \hat{y}_i)^2}{n} \right)^{1/2} \quad \dots(2)$$

Dimana :

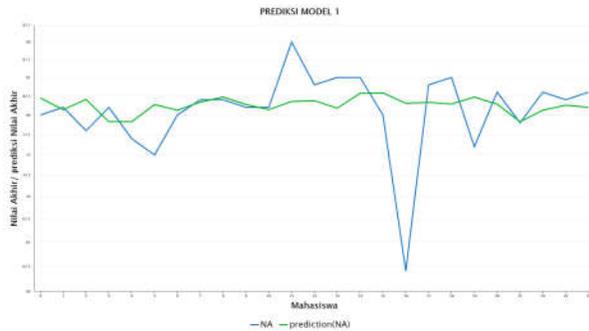
- y = nilai hasil observasi
- $\hat{y}$  = nilai hasil prediksi
- i = urutan data
- n = jumlah data

## III. HASIL DAN PEMBAHASAN

Model regresi linier berganda dibuat dalam beberapa model dengan prediktor yang berbeda untuk mendapatkan model terbaik dalam melakukan prediksi.

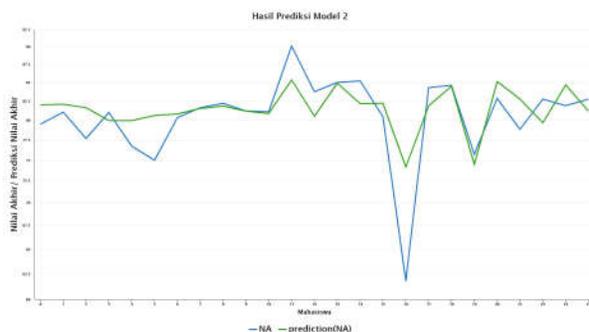
- Model #1: variabel prediktor adalah  $X_1, X_2$
- Model #2: variabel prediktor adalah  $X_1, X_2, X_3$
- Model #3: variabel prediktor adalah  $X_1, X_2, X_3, X_4$
- Model #4: variabel prediktor adalah  $X_2, X_3, X_4$

Pada model prediksi 1, prediktor yang digunakan adalah  $X_1$  dan  $X_2$ , yaitu nilai tugas dan nilai praktikum. Prediksi yang didapat dari model 1 ini menghasilkan RMSE sebesar  $5.004 \pm 0.000$ . Perbandingan prediksi nilai akhir dan nilai akhir yang sebenarnya seperti grafik pada gambar 2.



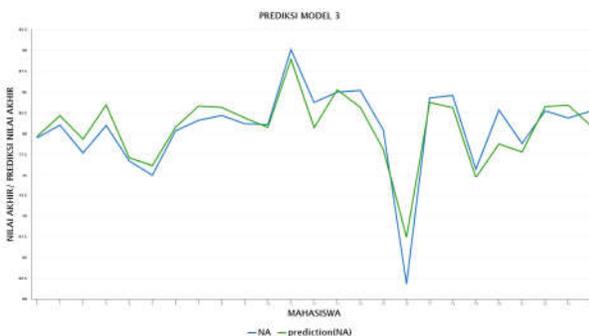
Gbr. 2 Grafik prediksi nilai akhir dan nilai akhir aktual dengan model 1

Sedangkan model prediksi kedua, prediktor yang digunakan adalah X1, X2 dan X3, yaitu 1 nilai tugas dan 2 nilai praktikum. Prediksi yang didapat lebih baik dari model 1 dengan nilai RMSE sebesar 3.538 +/- 0.000. Dan grafik perbandingan prediksi dan nilai akhir yang sebenarnya seperti pada gambar 3.



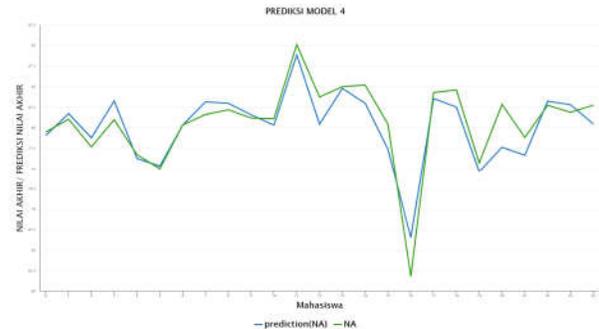
Gbr. 3 Grafik prediksi nilai akhir dan nilai akhir aktual dengan model 2

Untuk model prediksi ketiga, prediktor yang digunakan adalah X1, X2, X3 dan X4, yaitu 1 nilai tugas, 2 nilai praktikum dan nilai UTS. Prediksi yang didapat lebih baik dari model 1 dan model 2 dengan nilai RMSE sebesar 1.949 +/- 0.000. Grafik perbandingan prediksi dan nilai akhir yang sebenarnya seperti pada gambar 4.



Gbr. 4 Grafik prediksi nilai akhir dan nilai akhir aktual dengan model 3

Untuk model prediksi keempat, dicoba menggunakan prediktor yang digunakan adalah X2, X3 dan X4, yaitu hanya 2 nilai praktikum dan nilai UTS. Prediksi yang didapat lebih baik dari model 1 dan model 2 dengan nilai RMSE sebesar 2.043 +/- 0.000. Tetapi model 4 ini tidak lebih baik dari model 3. Perbandingan hasil prediksi dan nilai akhir yang sebenarnya seperti grafik pada gambar 5.



Gbr. 5 Grafik prediksi nilai akhir dan nilai akhir aktual dengan model 4

#### IV. KESIMPULAN

Memprediksi kinerja siswa akan meningkatkan hasil nilai mahasiswa dan memberikan pendekatan yang lebih baik kepada dosen untuk mengajar mahasiswa yang berisiko gagal. Model regresi, dibuat untuk membuat prediksi terbaik dengan akurasi tinggi. Dari 4 model yang dibuat, model 3 mempunyai nilai RMSE yang terbaik yaitu 1.949. Pada model 3 ini menggunakan predictor nilai tugas, nilai praktikum serta nilai UTS. Sehingga dapat diambil kesimpulan bahwa ketiga jenis nilai tersebut sangat berpengaruh terhadap kinerja mahasiswa pada pembelajaran matakuliah Pemrograman Dasar.

#### REFERENSI

- [1] P. Ajay Kumar, aurabh Pal "Data Mining Techniques in EDM for Predicting the Performance of Students" *International Journal of Computer and Information Technology*, Vol.2(06), November 2013.
- [2] Y. Surjeet Kumar & Saurabh Pal, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification", *World of Computer Science and Information Technology Journal*, Vol. 2, No. 2, hal. 51-56, March 2012
- [3] R.R.Rajalaxmi, P.Natesan, N.Krishnamoorthy, S.Ponni, "Regression Model for Predicting Engineering Students Academic Performance", *International Journal of Recent Technology and Engineering*, Vol.7 Issue-6S3, April, 2019.
- [4] F. Devine Grace D., "Predicting Student Academic Performance in Computer Organization Course: Using J48 Algorithm", *Indian Journal of Science and Technology*, Vol 11(47), hal. 1-8, December 2018.
- [5] IG.Sujatha, Sindhu.S & Savaridassan P, "Predicting Students Performance Using Personalized Analytics". *International Journal of Pure and Applied Mathematics*, Vol. 119 No. 12, hal. 229-238, 2018.
- [6] C. Huanyi and Paul A.S. Ward. "Predicting student performance using data from an Auto-grading system", *prod. CASCON '19*. ACM, 2019.
- [7] D. Oyerinde O., Chia P., "Predicting Students' Academic Performances - A Learning Analytics Approach using Multiple Linear Regression", *International Journal of Computer Applications*, Vol 57 - No 4, January 2017.
- [8] A. Ouafae El, Yasser El Alami El Madani, Lahcen Oughdir, Ahmed Dakkak, and Youssef El Alloui, "A Multiple Linear Regression-

Based Approach to Predict Student Performance”, *proc. AI2SD 2019*,  
hal.9-23.