

Estimasi Nilai Akhir Mata Pelajaran Komputer dan Jaringan Dasar Menggunakan Algoritma Linear Regression Berganda

¹M. Andrea Rossi,²Elinda,³Tisna Hidayat,⁴Ainunna'imah,⁵Tedy Setiadi
Informatika, Universitas Ahmad Dahlan^{1,2,3,4,5}

Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55191^{1,2,3,4,5}
m2000018327@webmail.uad.ac.id¹, elinda2000018033@webmail.uad.ac.id², tisna2000018313@webmail.uad.ac.id³,
ainun2000018016@webmail.uad.ac.id⁴, tedy.setiadi@tif.uad.ac.id⁵

Abstract - Vocational High School or SMK is a secondary level of education taken after completing from junior high school/equivalent. SMK plays a significant part in the preparation of human resources. The final grades of the students are difficult for teachers to predict. This estimated final grade is necessary for the teacher to evaluate the quality of the instruction given. In order to ascertain the association between the variable daily average values and competency scores on the final grades in Computer and Basic Networks subjects, a data mining strategy was utilized in conjunction with the estimate method utilizing the multiple linear regression algorithm. The data used is value data for one semester with a total of 134. In order to increase the quality of learning in the event that the grade does not satisfy the completeness requirements, it is anticipated that the Multiple Linear Regression algorithm will be able to estimate the final grades that are likely to be achieved. To determine the scope of the fault discovered, the RMSE (Root Mean Squared Error) is utilized.

Keyword: Mining, Estimate, Linear Regression

Abstrak – Sekolah Menengah Kejuruan atau SMK merupakan tingkatan pendidikan menengah yang ditempuh setelah lulus dari SMP atau Sekolah Menengah Pertama/ sederajat. SMK memiliki peranan penting untuk mempersiapkan Sumber Daya Manusia. Guru memiliki permasalahan dalam mengestimasi nilai akhir siswa. Estimasi nilai akhir ini diperlukan oleh guru dalam mengevaluasi terkait kualitas pembelajaran yang diberikan. Pada penelitian ini, digunakan pendekatan data mining dengan metode estimasi menggunakan algoritma Regresi Linear Berganda guna mengetahui hubungan antara variabel nilai rata-rata harian dan nilai kompetensi terhadap nilai akhir mata pelajaran Komputer dan Jaringan Dasar. Data yang digunakan adalah data nilai selama satu semester dengan jumlah sebanyak 134. Dengan menggunakan algoritma Regresi Linear Berganda diharapkan mampu memperkirakan hasil nilai akhir yang kemungkinan diperoleh sehingga dapat meningkatkan kualitas belajar jika nilai tidak memenuhi syarat ketuntasan. RMSE (Root Mean Squared Error) digunakan untuk melakukan pengecekan terhadap besarnya error yang didapat.

Keyword: Data Mining, Estimasi, Linear Regresi Berganda

I. PENDAHULUAN

Data mining merupakan teknik mencari informasi yang menarik atau pola pada data yang dipilih menggunakan metode tertentu. Pada penambahan data, metode, teknik atau algoritma sangat berbeda. Memilih algoritma atau metode yang tepat bergantung pada proses dan tujuan Penemuan Pengetahuan dalam Basis Data secara keseluruhan[1]. Berbagai macam metode yang ada pada data mining, salah satu contohnya adalah metode estimasi[2]. Metode estimasi berfungsi ketika dataset sampel data numerik dan memiliki judul. Selain itu, metode ini dapat digunakan untuk menebak nilai yang belum diketahui[3], misalnya untuk kasus saat ini nilai akhir mata pelajaran Komputer dan Jaringan Dasar.

Secara umum sistem pembelajaran lembaga pendidikan vokasional atau kejuruan terbagi menjadi pengajaran produktif, adaptif dan normatif. Komputer dan Jaringan dasar di SMK termasuk pembelajaran produktif, yaitu pembelajarannya berlangsung tidak di dalam kelas saja, tetapi juga di laboratorium, karena mata pelajaran ini merupakan materi praktikum sesuai dengan kurikulum SMK, pada yang menjadi dasar turunannya. Dalam kurikulum dan KI KD[4]. Selama satu semester, pada mata pelajaran ini tentunya dilakukan penilaian dari tugas-tugas yang diberikan. Penilaian-penilaian tersebut berupa penilaian pengetahuan maupun keterampilan. Untuk tugas-tugas terkait pengetahuan, akan didapatkan nilai rata-rata harian, sedangkan untuk keterampilan (praktik), akan didapatkan nilai kompetensi.

Pada penelitian ini menggunakan metode *Linear Regression* berganda. Metode ini mengukur data estimasi dengan garis lurus sebagai penggambaran dari hubungan korelasi antara dua atau lebih variabel. Estimasi *Linear regression* berganda digunakan untuk mempelajari hubungan setiap variabel pada proses perkiraan data[5]. Pendekatan ini dilakukan untuk

mengetahui pengaruh nilai-nilai penugasan, baik pengetahuan maupun keterampilan terhadap nilai akhir mata pelajaran komputer dan jaringan dasar.

Standar error hasil estimasi dihitung menggunakan RMSE (*Root Mean Square Error*). RMSE yaitu ukuran yang digunakan berdasarkan perbedaan antara nilai-nilai yang diprediksi oleh estimator atau model dan nilai-nilai yang benar-benar diamati. Analisis regresi mempunyai beberapa kegunaan diantaranya untuk mendeskripsikan fenomena data atau kasus yang diteliti, untuk tujuan kontrol, dan untuk tujuan perkiraan atau prediksi[6]. Dengan perhitungan estimasi dengan regresi linear berganda tersebut diharapkan mampu untuk memperkirakan hasil nilai akhir yang kemungkinan diperoleh dan mampu meningkatkan kualitas belajar jika nilai tidak memenuhi syarat ketuntasan.

Berdasarkan penelitian Amrin, S.Si, M.Kom dengan judul penelitian Data Mining Dengan Regresi Linear Berganda Untuk Peramalan Tingkat Inflasi menyatakan bahwa Algoritma Regresi Linear Berganda adalah analisis yang mempunyai variabel bebas lebih dari satu atau disebut juga dengan analisis regresi linear berganda. Teknik regresi linear berganda ini digunakan dalam mengetahui ada atau tidaknya pengaruh yang signifikan dari dua atau lebih variabel bebas ($X_1, X_2, X_3, \dots, X_n$) terhadap variabel terikat (Y)[7].

Berikut bentuk persamaan untuk regresi linear berganda:

$$Y = p + q_1X_1 + q_2X_2 + \dots + q_nX_n$$

Keterangan :

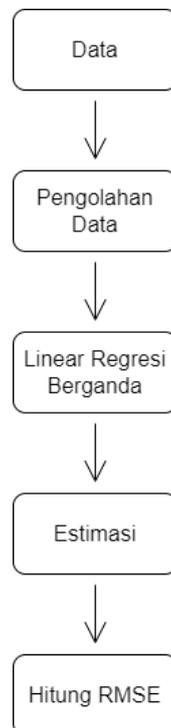
Y : Variabel dependen / terikat

p : Konstanta

q_1, q_n : Koefisien regresi

X_1, X_n : Variabel bebas

II. METODOLOGI PENELITIAN



Gambar 1. Langkah-Langkah Penelitian

Sesuai dengan **Gambar 1**, langkah-langkahnya sebagai berikut:

1. Dari data-data nilai mata pelajaran Komputer dan Jaringan Dasar, data yang digunakan berupa kumpulan nilai.
2. Data diolah dan menghilangkan atribut yang memiliki korelasi rendah terhadap atribut nilai akhir.
3. Kemudian tentukan variabel, perhitungan dan aturannya. Tuliskan persamaan regresi linear berganda.
4. Melakukan estimasi nilai akhir dari hasil perhitungan dengan persamaan regresi linear berganda. RMSE (*Root Mean Squared Error*) digunakan untuk melakukan pengecekan terhadap besarnya error yang didapat.

Analisis yang meneliti sejumlah besar data untuk menentukan hubungan yang sebelumnya belum diketahui dan dua metode baru yang digunakan untuk meringkas data sehingga lebih mudah dipahami dan digunakan ke pemilih data merupakan pengertian data mining. Data mining bukan merupakan bidang yang sepenuhnya baru. Fakta bahwa pada data mining tidak sedikit teknik dan aspek dari bidang ilmu yang telah ditetapkan sebelumnya adalah salah satu contoh kesulitan dalam pendefinisian data mining. Data mining digunakan untuk perbaikan terhadap teknik tradisional agar mampu menangani[8]:

1. Data dengan jumlah yang sangat besar.
2. Dimensi data yang tinggi.
3. Data-data yang berbeda (heterogen) dan sifat yang berbeda.

Tahap pertama yang dilakukan pada proses data mining, yaitu menyeleksi data dari sumber ke target, selanjutnya tahap preprocessing yang berguna untuk perbaikan kualitas data, data mining, dan tahapan yang terakhir yaitu interpretasi dan evaluasi yang akan menghasilkan keluaran berupa pengetahuan yang baru dan nantinya dapat berkontribusi lebih baik[3].

Regresi linear berganda

Linear Regression berganda memperluas regresi linear sederhana lebih dari satu variabel bebas (independen)[9]. Pada saat hasil atau kategori adalah numerik serta semua atribut adalah numerik, regresi linear merupakan teknik penyelesaian yang tepat. Ini merupakan metode statistik yang paling penting. Tujuannya ialah untuk mengekspresikan kelas sebagai kombinasi linear atribut dengan bobot yang sudah ditentukan[10]:

$$Y = p + p_1X_1 + p_2X_2 + \dots + p_kX_k$$

Dimana Y adalah kategorinya; X_1, X_2, \dots, X_k ialah nilai atributnya; p_1, p_2, \dots, p_k adalah bobotnya.

Bobot dapat dihitung dari data sampel. Di sinilah anotasi akan sedikit rumit, karena memerlukan cara untuk mengekspresikan nilai atribut untuk setiap sampelnya. Pada contoh pertama, misalkan ada kelas, katakanlah Y , dan nilai atribut X_1, X_2, \dots, X_k (2), di mana superskrip ialah contoh pertama.

Nilai prediksi kelas derajat pertama dapat ditulis sebagai :

$$p_n + p_1X_1 + p_2X_2 + \dots + p_kX_k = \sum_{i=0}^k p_iX_i$$

Estimasi

Estimasi ialah metode yang digunakan untuk menghitung nilai populasi berdasarkan nilai sampel[11]. Perkiraan biasanya dibutuhkan untuk membantu membuat keputusan yang lebih baik, merencanakan pekerjaan, dan menentukan seberapa lama sebuah proyek harus diselesaikan beserta biayanya, tentukan apakah proyek itu layak, kembangkan persyaratan arus kas, tentukan seberapa baik kemajuan proyek, buat anggaran bertahap waktu, serta tetapkan garis dasar proyek[10].

Root mean squared error (RMSE)

Root mean squared error atau biasa disingkat RMSE ialah metode alternatif evaluasi dari teknik perkiraan yang digunakan untuk mengukur ketepatan hasil perkiraan model[12]. RMSE adalah akar kuadrat rata-rata dari perbedaan antara nilai estimasi data dan nilai yang diamati. Semakin rendah nilai RMSE, semakin valid datanya[13]. Rumus RMSE sebagai berikut:

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(x_i - f_i)^2}}{n}$$

III. HASIL DAN PEMBAHASAN

Data dan business understanding

Dataset yang digunakan pada jurnal ini berisi nilai pelajaran komputer dan jaringan dasar selama satu semester. Dataset ini berisi 11 attributes, diantaranya:

1. No : merupakan no urut siswa
2. No. Induk : merupakan nomor identitas siswa
3. Nama : nama siswa
4. L/P : jenis kelamian siswa
5. Jumlah Hadir : jumlah kehadiran siswa selama satu semester
6. Rata2 Harian : nilai rata2 harian siswa satu semester
7. MID : nilai mid (ujian tengah semester) siswa
8. UAS : nilai uas (ujian akhir semester) siswa

- 9. NP : nilai pengetahuan siswa
- 10. NK : nilai kompetensi siswa
- 11. NR : nilai akhir siswa

Mencari nilai yang hilang pada dataset

```
# Memeriksa jumlah missing value pada dataset
raw_data.isnull().sum()

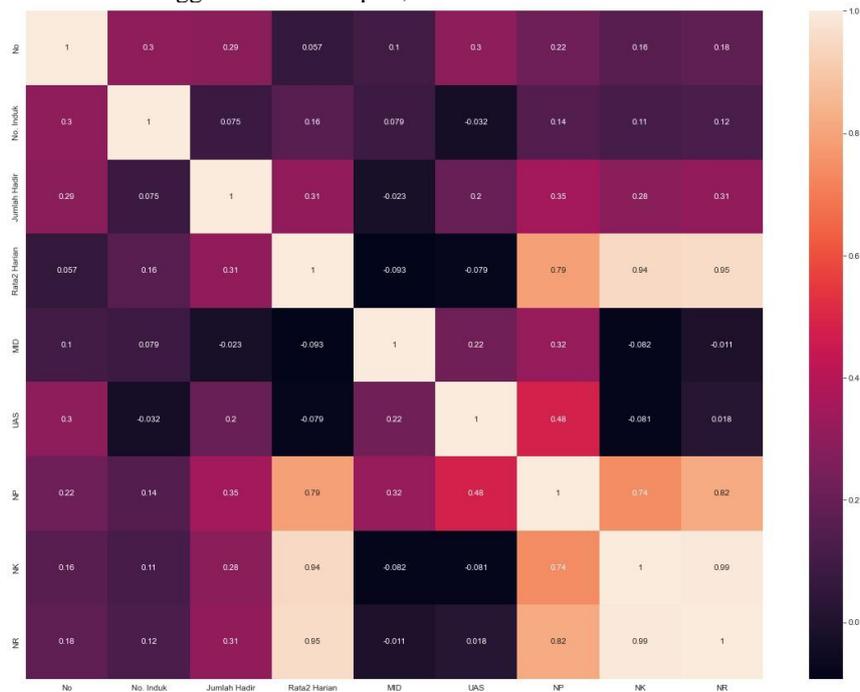
No                0
No. Induk        0
Nama              0
L/P              0
Jumlah Hadir     0
Rata2 Harian     0
MID              0
UAS              0
NP               0
NK               0
NR               0
dtype: int64
```

Gambar 2. Missing Values

Gambar 2 menunjukkan tidak adanya missing values pada dataset nilai pelajaran komputer dan jaringan dasar yang digunakan pada jurnal ini.

Mencari korelasi tiap variabel

Korelasi tiap variabel dicari menggunakan Matrixplot, berikut hasil korelasi antar variabel.



Gambar 3. Korelasi antar Variabel

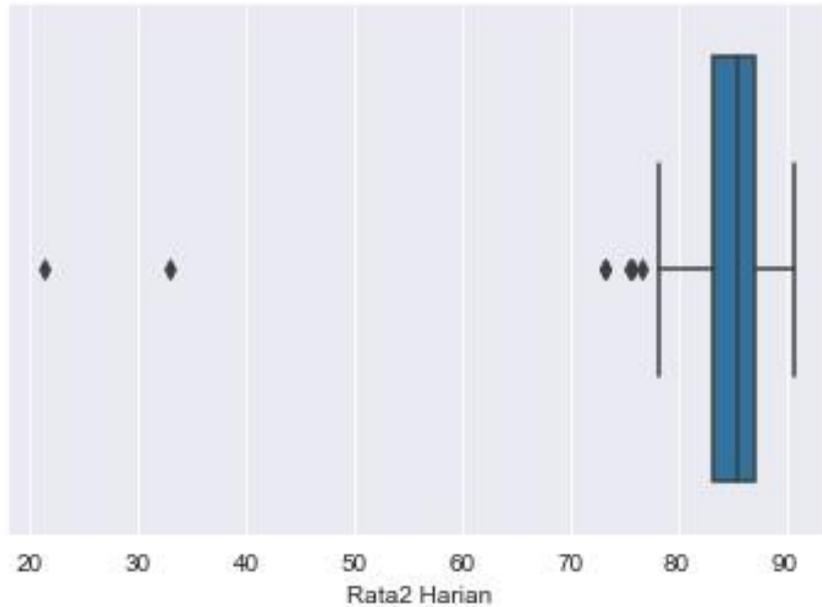
Dari Gambar 3, semakin gelap warnanya maka korelasinya semakin rendah. Variabel yang memiliki korelasi rendah terhadap nilai akhir akan dieliminasi atau dihilangkan, sehingga hanya menggunakan variabel rata-rata harian dan NK (Nilai Kompetensi) karena korelasinya terhadap variabel Nilai Akhir (NR) tinggi (mendekati 1).

```
# Drop id column
data = data1.drop(['No', 'No. Induk', 'Jumlah Hadir', 'MID', 'UAS', 'NP', 'Nama', 'L/P'], axis = 1)
target = data['NR']
data = data.drop(['NR'], axis = 1)
data.head()
```

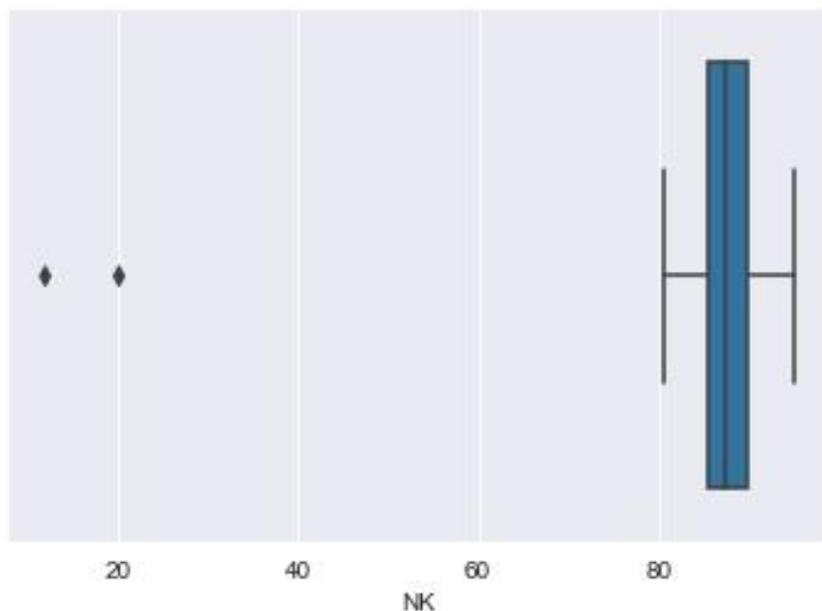
Gambar 4. Drop Variabel

Variabel yang tersisa pada dataset adalah Rata2 Harian dan NK, variabel NR di hilangkan karena akan menjadi target prediksi. Variabel-variabel yang dihilangkan dapat dilihat pada **Gambar 4**.

Menghilangkan outlier dari kumpulan data



Gambar 5. Outlier Rata-Rata Harian



Gambar 6. Outlier Nilai Kompetensi

Bintik hitam yang ada pada **Gambar 5** dan **Gambar 6** menandakan adanya *outlier* pada kumpulan data. Oleh karena itu, outlier tersebut harus dihilangkan terlebih dahulu sebelum digunakan.

```
dfr = raw_data['Rata2 Harian']
dfr_Q1 = dfr.quantile(0.25)
dfr_Q3 = dfr.quantile(0.75)
dfr_IQR = dfr_Q3 - dfr_Q1
dfr_lowerend = dfr_Q1 - (1.5 * dfr_IQR)
dfr_upperend = dfr_Q3 + (1.5 * dfr_IQR)
dfr_outliers = dfr[(dfr < dfr_lowerend) | (dfr > dfr_upperend)]
dfr_outliers
```

32	21.43
41	32.86
110	76.67
117	73.33
118	75.83
119	73.33
120	75.50

Name: Rata2 Harian, dtype: float64

Gambar 7. Nilai Outlier Rata-Rata Harian

Gambar 7 menampilkan nomor indeks outlier yang ada pada kumpulan data yakni pada nomor indeks ke-32, 41, 110, 117, 118, 119, 120. Nilai outlier tersebut yang berada pada nilai rata-rata harian.

```
dfk = raw_data['NK']
dfk_Q1 = dfk.quantile(0.25)
dfk_Q3 = dfk.quantile(0.75)
dfk_IQR = dfk_Q3 - dfk_Q1
dfk_lowerend = dfk_Q1 - (1.5 * dfk_IQR)
dfk_upperend = dfk_Q3 + (1.5 * dfk_IQR)
dfk_outliers = dfk[(dfk < dfk_lowerend) | (dfk > dfk_upperend)]
dfk_outliers
```

32	11.88
41	20.00

Name: NK, dtype: float64

Gambar 8. Nilai Outlier Nilai Kompetensi

Gambar 8 menampilkan nomor indeks outlier yang ada pada kumpulan data yakni pada nomor indeks ke-32, 41. Nilai outlier tersebut yang berada pada nilai kompetensi.

Entri data pada nomor indeks ke-32, 41, 110, 117, 118, 119, 120 akan dihapus atau dihilangkan karena mengandung outlier.

Menghitung persamaan linear

```
# Mengimport library LinierRegression
from sklearn.linear_model import LinearRegression

# Melakukan Linier Regression
model_LR = LinearRegression(normalize=False)
model_LR.fit(X_train, y_train)
print("num_ftrs = {}, num_coeff = {}".format(X_train.shape[1], len(model_LR.coef_)))
reg_coeff = dict(zip(data.columns, model_LR.coef_))
print('Model intercept: ', model_LR.intercept_)
print(reg_coeff)

num_ftrs = 2, num_coeff = 2
Model intercept: 4.3274582325509385
{'Rata2 Harian': 0.20932644927868618, 'NK': 0.7298694417711961}
```

Gambar 9. Koefisien Regresi

Pada **Gambar 9** didapatkan konstanta dan koefisien regresi, sehingga persamaan regresi linear untuk melakukan perhitungan dalam estimasi mencari nilai akhir dapat dituliskan sebagai berikut:

$$Y = 4.33 + 0.21X_1 + 0.73X_2$$

Persamaan di atas dapat digunakan untuk estimasi nilai akhir dengan memasukkan nilai X_1 (rata-rata harian) dan X_2 (nilai kompetensi).

Estimasi nilai akhir

Dengan menggunakan persamaan linear $Y = 4.33 + 0.21X_1 + 0.73X_2$ maka untuk memperhitungkan nilai akhir mata pelajaran Komputer dan Jaringan Dasar dapat dengan mudah dilakukan. Sebagai contoh, nilai $X_1 = 80$ dan $X_2 = 85$ maka perhitungannya sebagai berikut:

$$Y = 4.33 + 0.21 * 80 + 0.73 * 85$$

$$Y = 4.33 + 16.8 + 62.05$$

$$Y = 83.18$$

Dari hasil perhitungan di atas, estimasi nilai akhir yang didapatkan yaitu 83.18, dimana X_1 merupakan nilai rata-rata harian dan X_2 nilai kompetensi yang ada pada dataset.

Menghitung RMSE (Root Mean Squared Error)

```
print('RMSE:', np.sqrt(mean_squared_error(y_test, y_pred)))
```

```
RMSE: 0.7047423179518819
```

Gambar 10. Nilai RMSE

Pada penelitian ini, didapatkan hasil perhitungan *Root Mean Squared Error* sebesar 0.7047423179518819 yang dapat dilihat pada **Gambar 10**. Hasil tersebut tergolong kecil dengan jumlah data sebanyak 134. Semakin banyak jumlah data maka jumlah RMSE menjadi semakin sedikit. Seperti yang telah disebutkan bahwa RMSE digunakan untuk melakukan pengecekan terhadap besarnya error yang didapat. Dengan demikian, nilai RMSE yang kecil menunjukkan bahwa tingkat akurasi dari estimasi yang telah dilakukan terbilang tinggi.

IV. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan dalam jurnal ini, dapat disimpulkan bahwa derajat akhir Dasar Komputer dan Jaringan dapat diestimasi dengan algoritma regresi linier berganda. Analisis ini menunjukkan bahwa terdapat korelasi yang signifikan antara nilai akhir dengan rata-rata harian dan nilai kompetensi pada mata pelajaran ini. Hasil regresi menunjukkan bahwa model yang dibuat dapat memprediksi skor akhir dengan tingkat akurasi yang cukup tinggi.

Hasil analisis menunjukkan bahwa kompetensi mata pelajaran memiliki pengaruh terbesar terhadap hasil akhir, diikuti oleh rata-rata harian siswa. Hal ini menunjukkan bahwa jika siswa ingin mendapatkan nilai yang baik, mereka perlu belajar dengan giat dan juga mendapatkan nilai yang baik dalam ujian tertulis dan ujian praktek. Dalam konteks pengajaran, guru dapat menggunakan hasil ini untuk menentukan fokus pengajaran dan mengevaluasi keefektifan metode pengajaran yang digunakan. Guru dapat menggunakan templat ini untuk mengevaluasi nilai akhir siswa dan membuat saran yang sesuai untuk membantu siswa meningkatkan nilai akhir mereka.

Secara umum, algoritma regresi linier berganda dapat digunakan sebagai metode untuk mengestimasi nilai akhir dari topik dasar komputasi dan jaringan. Perhatikan, bagaimanapun, bahwa hasil analisis ini hanya berlaku untuk data yang digunakan dalam jurnal ini dan mungkin tidak berlaku untuk data lainnya.

REFERENSI

- [1] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perkam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [2] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," J. MEDIA Inform. BUDIDARMA, vol. 4, no. 3, p. 855, Jul. 2020, doi: 10.30865/mib.v4i3.2312.
- [3] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," 2021. [Online]. Available: <https://hostjournals.com/>
- [4] I. D. P. Nasional, "Kamus Besar Bahasa Indonesia: Pusat Bahasa," 2008.
- [5] R. Gunawan, N. B. Nugroho, and R. Arbiyanto, "Penerapan Data Mining Untuk Estimasi Laju Pertumbuhan Produk Domestik Regional Bruto (PDRB) Perkapita Atas Dasar Harga Berlaku Menurut Lapangan Usaha Pada Kota

- Medan Menggunakan Metode Regresi Linier Berganda,” J. Teknol. Sist. Inf. dan Sist. Komput. TGD, vol. 1, no. 2, pp. 106–113, 2018.
- [6] A. Fikri, “Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression,” Univ. Dian Nuswantoro, Semarang, 2013.
- [7] A. Amrin, “Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi,” Techno Nusa Mandiri J. Comput. Inf. Technol., vol. 13, no. 1, pp. 74–79, 2016.
- [8] D. H. Kamagi and S. Hansun, “Implementasi Data Mining dengan Algoritma C4. 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa,” Ultim. J. Tek. Inform., vol. 6, no. 1, pp. 15–20, 2014.
- [9] M. Tranmer, J. Murphy, M. Elliot, and M. Pampaka, “Multiple Linear Regression (2 nd Edition),” 2020. [Online]. Available: <https://hummedia.manchester.ac.uk/institutes/cmist/a>
- [10] E. I. A. Warih and Y. Rahayu, “Penerapan Data Mining Untuk Menentukan Estimasi Produktivitas Tanaman Tebu Dengan Menggunakan Algoritma Linier Regresi Berganda Di Kabupaten Rembang,” Tek. Inform. Fak. Ilmu Komputer, Univ. Dian Nuswantoro, Semarang, 2015.
- [11] A. Z. Siregar, “Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru,” Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen), vol. 2, no. 3, pp. 133–137, 2021.
- [12] L. Wiranda, M. Sadikin, J. T. Informatika, and F. I. Komputer, “Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT,” Metiska Farma. Jur. Tek. Inform. Fak. Ilmu Komput., 2019.
- [13] F. Febrianti, M. Hafiyusholeh, and A. H. Asyhar, “Perbandingan Pengklusteran data iris menggunakan metode k-means dan fuzzy c-means,” J. Mat. MANTIK, vol. 2, no. 1, pp. 7–13, 2016.