Praktikum 3: Machine Learning – Regresi Dan Evaluasi Model

**Muhammad Shiddiq 1 - 0110222199 1**

1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: muhammadshiddiq785@gmail.com

**Abstract.** Regresi merupakan salah satu teknik dalam pembelajaran mesin dan statistika yang digunakan untuk memodelkan hubungan antara variabel independen (prediktor) dan variabel dependen (target). Tujuan utamanya adalah untuk memprediksi nilai output berdasarkan input yang diberikan dengan membangun fungsi terbaik yang menggambarkan pola data. Proses regresi dapat dilakukan menggunakan berbagai metode seperti regresi linear, regresi polinomial, dan regresi non-linear lainnya. Setelah model regresi dibangun, tahap evaluasi model dilakukan untuk menilai seberapa baik model tersebut mampu memprediksi data. Evaluasi biasanya menggunakan metrik seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R-squared (R²). Metrik-metrik ini membantu menentukan akurasi, efisiensi, serta kemampuan generalisasi model terhadap data baru, sehingga model yang dihasilkan tidak hanya cocok pada data pelatihan tetapi juga memiliki performa baik pada data uji.

1. Praktikum Mandiri – Membuat Model Prediksi Jumlah Penyewaan Sepeda

Model prediksi jumlah penyewaan sepeda dibuat untuk memperkirakan berapa banyak sepeda yang akan disewa pada waktu tertentu, misalnya per jam atau per hari. Tujuan utamanya adalah membantu pengelola layanan sepeda agar dapat mengatur jumlah sepeda yang tersedia sesuai dengan kebutuhan pengguna. Dalam pembuatannya, data historis penyewaan sepeda digunakan, yang biasanya berisi informasi seperti tanggal, suhu, kelembapan, kecepatan angin, dan kondisi cuaca.

Metode yang umum digunakan untuk membuat model ini adalah regresi, karena metode ini dapat menggambarkan hubungan antara faktor-faktor tersebut dengan jumlah sepeda yang disewa. Setelah model dibangun, langkah selanjutnya adalah melakukan evaluasi untuk mengetahui seberapa baik model tersebut dalam memprediksi data baru. Beberapa ukuran yang digunakan antara lain MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), dan R² (R-squared). Nilai MAE dan RMSE yang kecil menunjukkan bahwa prediksi model cukup akurat, sedangkan nilai R² yang tinggi menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data. Dengan hasil ini, model dapat digunakan sebagai dasar pengambilan keputusan untuk mengoptimalkan penyediaan sepeda di berbagai kondisi.

**1.1 Membuat Folder**

Langkah pertama kita harus membuat folder yang terstruktur dan juga rapih di google drive.

|  |
| --- |
|  |
| **Gambar 1.** Membuat folder di google drive, agar mudah untuk diakses |

***1.2 Membuat file notebook google colab***

Selanjutnya membuat file notebook di google colab untuk praktikum.

|  |
| --- |
|  |
| **Gambar 2.** Membuat file google colab |

**1.3 *Menghubungkan google colab dengan google drive***

Selanjutnya menghubungkan google colab dengan google drive menggunakan perintah

“From google.colab import drive

Drive,mount(‘/content/drive’)”.

|  |
| --- |
|  |
| **Gambar 3.** Menghubungkan google colab dengan google drive |

***1.4******Meng install pandas***

Selanjutnya meng install library pandas dengan perintah “!pip install pandas”.

|  |
| --- |
|  |
| **Gambar 4.** Meng install pandas. |

***1.5******Meng import library pandas***

Selanjutnya meng import library pandas dengan perintah “import pandas as pd”. Pandas adalah perpustakaan Python sumber terbuka yang banyak digunakan untuk analisis dan manipulasi data. Perpustakaan ini menyediakan struktur data yang kuat dan fleksibel, terutama Series dan DataFrame, yang dirancang untuk menangani data terstruktur dengan efisien.

|  |
| --- |
|  |
| **Gambar 5.** Mengimport library pandas |

***1.6******Membaca dataset***

Selanjutnya membaca dataset day.csv yang ada di google drive menggunakan perintah

“df = pd.read\_csv('/content/drive/MyDrive/MACHINELEARNING/PRAKTIKUM/PRAKTIKUM3/DATA/day.csv')

df”

|  |
| --- |
|  |
| **Gambar 6.** Membaca dataset day.csv. |
| **Tabel 1.** Berikut adalah hasil dataset yang telah dibaca. |

***1.7******Mengecek informasi dataset***

Selanjutnya mengecek informasi dataset yang dibaca, dari total, jumlah kolum, missing value, dan type data menggunakan perintah

“df.info()”

|  |
| --- |
|  |
| **Gambar 7.** Mengecek informasi dataset. |

***1.8 Mencari nilai statistik deskriptif secara cepat***

|  |
| --- |
| Selanjutnya mencari nilai statistik dari dataset dengan cepat menggunakan perintah, df.describe(). |
| **Gambar 8.** Mencari nilai statistik deskriptif. |

***1.9 Menentukan variable independent dan dependent***

|  |
| --- |
| Selanjutnya menentukan variable independent dan dependent. |
| **Gambar 9.** Mencari nilai korelasi. |

***1.10 Membagi data testing dan training***

|  |
| --- |
|  |
| **Gambar 10.** Membagi data. |
|  |

**1.11 Menginstall model linear regresi**

|  |
| --- |
|  |
| **Gambar 11.** Menginstall model linear regresi. |
|  |

**1.12 Menyiapkan model prediksi**

|  |
| --- |
|  |
| **Gambar 12.** Menyiapkan model prediksi. |

**1.13 Evaluasi model**

|  |
| --- |
|  |
| **Gambar 13.** Mengevaluasi model. |

**1.14 Visualisasi Data**

|  |
| --- |
|  |
| **Gambar 14.** Hasil Visualisasi. |

***Link Github :*** [***https://github.com/Shid2iq/Machine-Learning***](https://github.com/Shid2iq/Machine-Learning)

Referensi:

Munir, S., Seminar, K. B., Sudradjat, Sukoco, H., & Buono, A. (2022). The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery. *Information*, *14*(1), 10. https://doi.org/10.3390/info14010010

Seminar, K. B., Imantho, H., Sudradjat, Yahya, S., Munir, S., Kaliana, I., Mei Haryadi, F., Noor Baroroh, A., Supriyanto, Handoyo, G. C., Kurnia Wijayanto, A., Ijang Wahyudin, C., Liyantono, Budiman, R., Bakir Pasaman, A., Rusiawan, D., & Sulastri. (2024). PreciPalm: An Intelligent System for Calculating Macronutrient Status and Fertilizer Recommendations for Oil Palm on Mineral Soils Based on a Precision Agriculture Approach. *Scientific World Journal*, *2024*(1). https://doi.org/10.1155/2024/1788726