Praktikum 4: Machine Learning – Logist

**Muhammad Shiddiq 1 - 0110222199 1**

1 Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: muhammadshiddiq785@gmail.com

**Abstract.** Logistic Regression adalah salah satu algoritma machine learning yang digunakan untuk memprediksi variabel kategori (klasifikasi) berdasarkan satu atau lebih variabel input. Berbeda dengan regresi linear yang menghasilkan nilai kontinu, logistic regression menghasilkan probabilitas suatu data termasuk ke dalam kelas tertentu, biasanya antara 0 dan 1. Model ini menggunakan fungsi sigmoid (logit function) untuk mengubah output linear menjadi nilai probabilitas. Logistic regression sering digunakan dalam berbagai kasus seperti prediksi penyakit (ya/tidak), analisis kelulusan siswa (lulus/tidak), atau deteksi spam (spam/tidak spam). Selain sederhana dan mudah diinterpretasikan, model ini juga menjadi dasar bagi banyak metode klasifikasi lain yang lebih kompleks.

1. Praktikum Mandiri – Membuat Model Logistik Regresi Calon Pembeli Mobil

Model Logistic Regression Calon Pembeli Mobil dibuat untuk memprediksi apakah seseorang berpotensi membeli mobil atau tidak berdasarkan berbagai faktor yang memengaruhi keputusan pembelian. Tujuan utamanya adalah membantu pihak dealer atau perusahaan otomotif dalam mengidentifikasi calon pembeli potensial, sehingga strategi pemasaran dan penawaran dapat dilakukan dengan lebih tepat sasaran. Dalam pembuatannya, digunakan data historis calon pembeli, yang biasanya berisi informasi seperti usia, pendapatan, status pernikahan, pekerjaan, preferensi kendaraan, serta riwayat kredit.

Metode logistic regression digunakan karena mampu memodelkan hubungan antara faktor-faktor tersebut dengan probabilitas seseorang membeli mobil (ya/tidak). Setelah model dibangun, dilakukan evaluasi kinerja untuk mengetahui seberapa baik model dalam melakukan klasifikasi terhadap data baru. Beberapa ukuran yang digunakan antara lain accuracy (akurasi), precision, recall, dan F1-score. Nilai akurasi dan F1-score yang tinggi menunjukkan bahwa model memiliki kemampuan prediksi yang baik dalam membedakan calon pembeli dan bukan pembeli. Dengan hasil ini, model dapat dimanfaatkan untuk mendukung pengambilan keputusan bisnis, seperti menentukan target promosi, strategi penjualan, dan efisiensi anggaran pemasaran.

**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 yang akan digunakan***

Langkah pertama dalam praktikum ini adalah melakukan import terhadap seluruh library yang dibutuhkan.

|  |
| --- |
|  |
| **Gambar 5.** Mengimport library  **Tabel 1.** Penjelasan Library |

***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/calonpembelimobil.csv')

df”

|  |
| --- |
|  |
| **Gambar 6.** Membaca dataset calonpembelimobil.csv. |
| **Tabel 2.** 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 Mengecek missing value***

|  |
| --- |
| Selanjutnya mencari missing value pada data dengan perintah df.isnull().sum() |
| **Gambar 8.** Mengecek missing value |

***1.9 Mengecek data duplikat***

|  |
| --- |
| Selanjutnya mengecek data duplikat dengan perintah df.duplicated().sum() |
| **Gambar 9.** Mengecek duplicate |

***1.10 Mengubah nilai kolum menjadi unik***

Selanjutnya mengubah nilai kolum status dan kelamin menjadi unik dengan perintah print(df['Status'].unique())

|  |
| --- |
|  |
| **Gambar 10.** Mengubah nilai kolum |
|  |

**1.11 *Mamping kolom kategori***

|  |
| --- |
|  |
| **Gambar 11.** Mamping kolom kategori. |
|  |

**1.12 *Mencari nilai korelasi***

|  |
| --- |
|  |
| **Gambar 12.** Mencari nilai korelasi |

|  |
| --- |
| **Tabel 3.** Mencari nilai korelasi |
|  | |
|  | |

**1.13 Visualisasi korelasi**

|  |
| --- |
|  |
| **Gambar 13.** Hasil Visualisasi. |
|  |

**1.14 *Menentukan fitur dan target***

|  |
| --- |
|  |
| **Gambar 14.** Menentukan fitur dan target. |

**1.15 *Membagi data menjadi training dan testing***

|  |
| --- |
|  |
| **Gambar 15.** Membagi data. |

**1.16 *Menginstall model logistic regression***

|  |
| --- |
|  |
| **Gambar 16.** Menginstall model. |
|  |

**1.17 *Membangun model logistic regression***

|  |
| --- |
|  |
| **Gambar 17.** Membangun model. |

**1.18 *Memprediksi dan mengevaluasi model***

|  |
| --- |
|  |
| **Gambar 18.** Memprediksi dan mengevaluasi. |

**1.19 *Visualisasi hasil evaluasi***

|  |
| --- |
|  |
| **Gambar 19.** Visualisasi hasil evaluasi.   |  |  | | --- | --- | |  |  | |

**1.20 *Clasiffication report***

|  |
| --- |
|  |
| **Gambar 20.** Classification report. |

**1.20 *Clasiffication report***

|  |
| --- |
|  |
| **Gambar 20.** Classification report. |

**1.21 *Interpretasi model logistic regression***

|  |
| --- |
|  |
| **Gambar 21.** Interpretasi model. |

**1.22 *Membuat data baru untuk menguji model***

|  |
| --- |
|  |
| **Gambar 22.** Membuat data baru |

***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