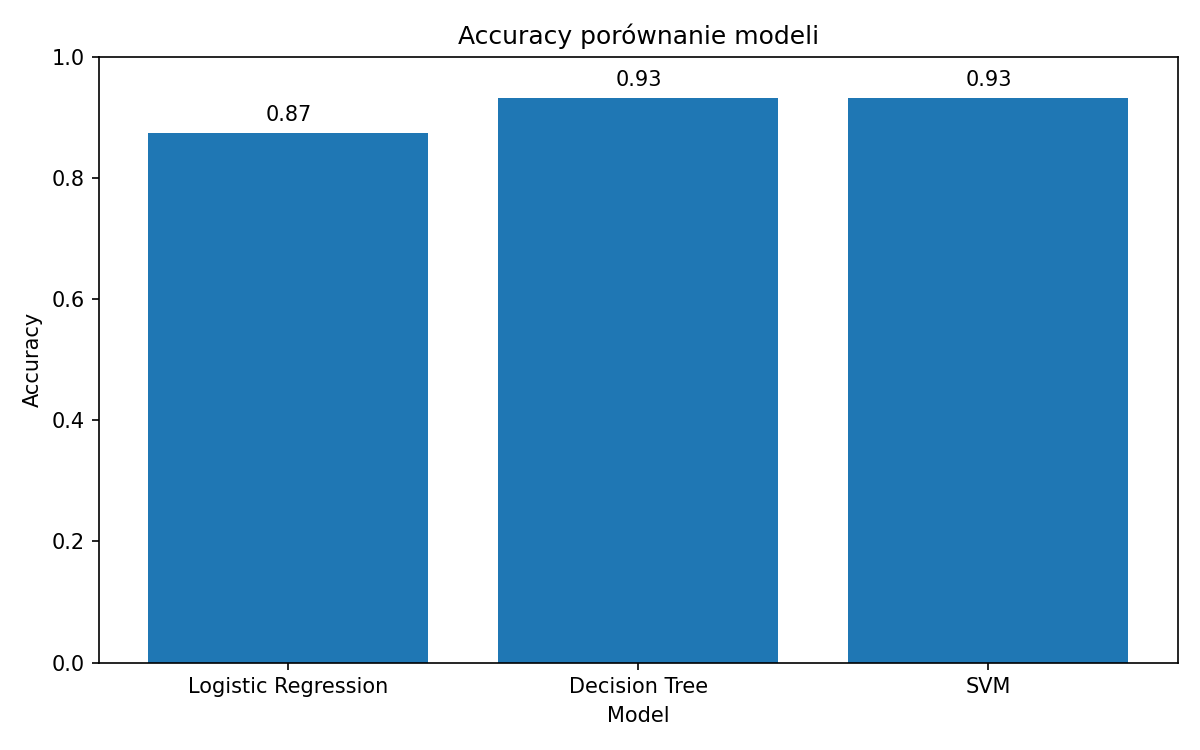
*Etap 2 Jakub P.*

**Zakres na 3.0:**  
„train.py” zawiera rozwiązanie, a plik iwynikowe w tym wykresy zapisywane są do folderu /results. Kolejne kroki wykonania:

* Wczytanie danych z pliku CSV do struktury DataFrame.
* Zbudowanie pipeline'u przetwarzania danych z wykorzystaniem Pipeline i ColumnTransformer.
* Zastosowanie imputacji wartości brakujących oraz kodowania cech kategorycznych (SimpleImputer, OneHotEncoder).
* Wybranie trzech modeli klasyfikacyjnych: regresji, drzewa decyzyjnego i SVM.
* Podział danych na zbiory treningowy i testowy z uwzględnieniem proporcji klas (stratify).
* Przeprowadzenie walidacji krzyżowej, trenowania modeli i predykcji.
* Wykonanie ewaluacji za pomocą raportu klasyfikacji i macierzy pomyłek.
* Zapisanie wyników i wizualizacji do plików.

Otrzymane wyniki:

| **Model** | **Accuracy** | **F1 (Macro)** |
| --- | --- | --- |
| Logistic Regression | 0.87 | 0.87 |
| Decision Tree | 0.93 | 0.93 |
| SVM | 0.93 | 0.93 |



**Logistic Regression**

**Cross-Validation Mean**: 0.8791

**Classification Report:**

| **Category** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Insufficient\_Weight | 0.87 | 1.00 | 0.93 | 54 |
| Normal\_Weight | 0.85 | 0.69 | 0.76 | 58 |
| Obesity\_Type\_I | 0.94 | 0.91 | 0.93 | 70 |
| Obesity\_Type\_II | 0.95 | 0.98 | 0.97 | 60 |
| Obesity\_Type\_III | 0.98 | 1.00 | 0.99 | 65 |
| Overweight\_Level\_I | 0.70 | 0.78 | 0.74 | 58 |
| Overweight\_Level\_II | 0.80 | 0.74 | 0.77 | 58 |
| **Weighted Average** | **0.88** | **0.87** | **0.87** | 423 |

**Accuracy**: 0.87

Obraz zawierający tekst, zrzut ekranu, diagram, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Decision Tree**

**Cross-Validation Mean**: 0.9236

**Classification Report:**

| **Category** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Insufficient\_Weight | 0.96 | 0.94 | 0.95 | 54 |
| Normal\_Weight | 0.85 | 0.88 | 0.86 | 58 |
| Obesity\_Type\_I | 0.93 | 0.94 | 0.94 | 70 |
| Obesity\_Type\_II | 0.97 | 0.98 | 0.98 | 60 |
| Obesity\_Type\_III | 1.00 | 1.00 | 1.00 | 65 |
| Overweight\_Level\_I | 0.89 | 0.84 | 0.87 | 58 |
| Overweight\_Level\_II | 0.91 | 0.91 | 0.91 | 58 |
| **Weighted Average** | **0.93** | **0.93** | **0.93** | 423 |

**Accuracy**: 0.93

Obraz zawierający tekst, zrzut ekranu, diagram, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**SVM**

**Cross-Validation Mean**: 0.9194

**Classification Report:**

| **Category** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Insufficient\_Weight | 0.95 | 0.96 | 0.95 | 54 |
| Normal\_Weight | 0.89 | 0.84 | 0.87 | 58 |
| Obesity\_Type\_I | 0.97 | 0.94 | 0.96 | 70 |
| Obesity\_Type\_II | 0.97 | 0.98 | 0.98 | 60 |
| Obesity\_Type\_III | 1.00 | 1.00 | 1.00 | 65 |
| Overweight\_Level\_I | 0.83 | 0.86 | 0.85 | 58 |
| Overweight\_Level\_II | 0.90 | 0.91 | 0.91 | 58 |
| **Weighted Average** | **0.93** | **0.93** | **0.93** | 423 |

**Accuracy**: 0.93

Obraz zawierający tekst, zrzut ekranu, diagram, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Zakres na 4.0 (5.0):**

Implementacje znajdują się w skryptach: „line\_reg” (1.a. + 3.) oraz „line\_logist”(2.a.b.+3.) oba z uwzględnieniem uwag 4. i 5. Wyniki są w folderach /regression\_results oraz /logistic\_results

**1. b.**   
Ograniczenia zamkniętej formuły regresji liniowej:

* Metoda nie jest skalowalna – dla dużych zbiorów danych odwracanie macierzy cech staje się obliczeniowo kosztowne.
* Wrażliwa na wielokolinearność – w przypadku silnie skorelowanych cech macierz może być osobliwa lub źle uwarunkowana.
* Ograniczona elastyczność – trudności w stosowaniu regularizacji, co ogranicza możliwości poprawy jakości modelu.
* Z tych powodów w praktyce preferowane są podejścia iteracyjne, takie jak metoda gradientu prostego.

**1. a.**• Wczytanie danych z pliku CSV i przygotowanie cech numerycznych oraz kategorycznych.

• Zakodowanie zmiennych kategorycznych za pomocą kodowania OneHot (get\_dummies).

• Podział danych na zbiór treningowy i testowy.

• Imputacja braków (SimpleImputer) i standaryzacja cech (StandardScaler).

• Dodanie biasu do macierzy cech (kolumna jedynek).

• Wybranie zmiennej ciągłej (Weight) jako cel (ground truth) w zadaniu regresyjnym.

Obraz zawierający Czcionka, biały, typografia, kaligrafia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.• Zaimplementowanie regresji liniowej z wykorzystaniem wzoru zamkniętego:

• Obliczenie predykcji na zbiorze testowym i ewaluacja modelu (MSE, RMSE, MAE, R²).

• Zapis wyników, wag i wizualizacji do plików (CSV, PNG).

• Porównanie wyników z modelem LinearRegression z biblioteki scikit-learn.

Otrzymane wyniki:  
Wyniki kilku predykcji zestawione z wartością właściwą:

| **Nr** | **Wartość rzeczywista** | **Wartość przewidywana** |
| --- | --- | --- |
| 1 | 133.85 | 128.75 |
| 2 | 89.99 | 91.14 |
| 3 | 69.00 | 67.57 |
| 4 | 80.00 | 80.01 |
| 5 | 105.06 | 109.99 |
| 6 | 45.00 | 41.90 |
| 7 | 85.30 | 86.57 |
| 8 | 59.26 | 65.67 |
| 9 | 114.53 | 116.66 |
| 10 | 58.50 | 56.32 |
| 11 | 80.00 | 78.18 |
| 12 | 120.92 | 117.70 |
| 13 | 147.30 | 128.27 |
| 14 | 80.00 | 86.21 |
| 15 | 102.78 | 106.65 |
| 16 | 103.18 | 106.43 |

Wykresy:  
Obraz zawierający diagram, Wykres, linia, tekst

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający linia, Wykres, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, diagram, linia, Wykres

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

2.a.b.

* Wczytanie danych z pliku CSV i przygotowanie cech numerycznych oraz kategorycznych.
* Zakodowanie zmiennych kategorycznych za pomocą kodowania OneHot (OneHotEncoder).
* Podział danych na zbiór treningowy i testowy z uwzględnieniem proporcji klas (stratify).
* Imputacja braków (SimpleImputer) i standaryzacja cech (StandardScaler).
* Implementacja regresji logistycznej z wykorzystaniem metody gradientu prostego.
* Funkcja softmax (jako uogólnienie regresji logistycznej dla przypadku wieloklasowego) została zaimplementowana z uwzględnieniem stabilizacji numerycznej.
* Obliczenie funkcji kosztu typu cross-entropy (CE) dla klasyfikacji.
* Iteracyjne aktualizowanie wag na podstawie obliczonego gradientu błędu w mini-batchach.
* Trening modelu przy użyciu wybranych hiperparametrów (np. współczynnik uczenia, liczba epok, wielkość batcha).
* Ewaluacja modelu na zbiorze testowym, obliczenie dokładności (accuracy), macierzy pomyłek, oraz cross-entropy loss.
* Zapisanie wyników do plików CSV i wizualizacji (accuracy, koszt).
* Porównanie wyników z klasycznym modelem regresji logistycznej z biblioteki scikit-learn.

Wyniki:  
**== Własna Softmax Regresja ==**

* **Accuracy**: 0.9054
* **Cross-Entropy Loss**: 0.3215

| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Insufficient\_Weight | 0.86 | 1.00 | 0.93 | 56 |
| Normal\_Weight | 0.94 | 0.74 | 0.83 | 62 |
| Obesity\_Type\_I | 0.96 | 0.92 | 0.94 | 78 |
| Obesity\_Type\_II | 0.92 | 0.97 | 0.94 | 58 |
| Obesity\_Type\_III | 1.00 | 1.00 | 1.00 | 63 |
| Overweight\_Level\_I | 0.83 | 0.79 | 0.81 | 56 |
| Overweight\_Level\_II | 0.81 | 0.92 | 0.86 | 50 |

**Overall:**

* **Accuracy**: 0.91
* **Macro avg**: Precision 0.90, Recall 0.91, F1-score 0.90
* **Weighted avg**: Precision 0.91, Recall 0.91, F1-score 0.90

**== sklearn LogisticRegression ==**

* **Accuracy**: 0.8652
* **Cross-Entropy Loss (log\_loss)**: 0.4127

| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| Insufficient\_Weight | 0.86 | 1.00 | 0.93 | 56 |
| Normal\_Weight | 0.89 | 0.63 | 0.74 | 62 |
| Obesity\_Type\_I | 0.93 | 0.88 | 0.91 | 78 |
| Obesity\_Type\_II | 0.89 | 0.97 | 0.93 | 58 |
| Obesity\_Type\_III | 1.00 | 1.00 | 1.00 | 63 |
| Overweight\_Level\_I | 0.72 | 0.73 | 0.73 | 56 |
| Overweight\_Level\_II | 0.74 | 0.84 | 0.79 | 50 |

**Overall:**

* **Accuracy**: 0.87
* **Macro avg**: Precision 0.86, Recall 0.86, F1-score 0.86
* **Weighted avg**: Precision 0.87, Recall 0.87, F1-score 0.86

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.  
Macierz konfuzji dla mojej implementacji:  
Obraz zawierający tekst, zrzut ekranu, diagram, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Dla scikit-learn:  
Obraz zawierający tekst, zrzut ekranu, diagram, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.