|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TOMASZ CHMIEL | |  | 2024 | PSI | | | | 1 | |
|  | **AUTOR** | **ROK AKAD.** |  | | | **SPEC.** |  | **GRUPA** |
| Wyjaśnialna sztuczna inteligencja. Algorytm LIME. | | | | |  | | | 2 | |
| **TEMAT** | | | **NR SPRAWOZDANIA** | |
| 13.12.2024 | Laboratorium maszynowej analizy danych | | | | | Anna Czaderna-Lekka | | | |
| **DATA WYKONANIA** | **PRZEDMIOT** | | | | | **PROWADZĄCY** | | | |

# Cel

Celem projektu jest zbudowanie modelu klasyfikacyjnego, który przewiduje, czy pasażer Titanica przeżył katastrofę na podstawie określonych cech. Projekt obejmuje następujące etapy:

1. **Analiza eksploracyjna danych (EDA)**:
   * Sprawdzenie kompletności danych i identyfikacja brakujących wartości.
   * Obliczenie podstawowych statystyk dla cech ilościowych (wiek, cena biletu) oraz wizualizacja rozkładu cech kategorycznych (płeć, klasa podróży).
   * Zbadanie zależności pomiędzy przeżyciem pasażera a wybranymi cechami (np. płeć, klasa podróży, wiek).
2. **Przetwarzanie danych**:
   * Uzupełnienie brakujących danych, np. w przypadku braku wartości w kolumnie 'Age' (wypełnienie medianą).
   * Usunięcie kolumn, które nie są istotne dla przewidywania (np. 'Cabin', 'Name', 'Ticket').
   * Przekształcenie cech kategorycznych na zmienne numeryczne za pomocą kodowania (np. płeć, port załadunku).
3. **Budowa modelu**:
   * Zbudowanie modelu klasyfikacyjnego przy użyciu algorytmu Random Forest.
   * Trenowanie modelu na danych treningowych i dokonywanie predykcji na danych testowych.
4. **Ewaluacja modelu**:
   * Ocena skuteczności modelu za pomocą raportu klasyfikacji oraz macierzy błędów.
   * Wizualizacja wyników za pomocą wykresu macierzy błędów oraz interpretacja wyników za pomocą algorytmu LIME.

# Materiały i metody

**Dane:**

Do realizacji zadania wykorzystano zbiór danych **"Titanic: Machine Learning from Disaster"**, dostępny w publicznym repozytorium Kaggle. Zbiór ten zawiera informacje o pasażerach Titanica oraz ich cechach demograficznych i statusie przeżycia w katastrofie. Zbiór zawiera 891 próbek, z których każda dotyczy jednego pasażera. Zbiór danych zawiera następujące cechy:

1. **PassengerId**: numer identyfikacyjny pasażera,
2. **Survived**: zmienna binarna, określająca, czy pasażer przeżył katastrofę (1) czy nie (0),
3. **Pclass**: zmienna określająca klasę podróży pasażera (1, 2, 3),
4. **Name**: imię i nazwisko pasażera,
5. **Sex**: płeć pasażera
6. **Age**: wiek pasażera,
7. **SibSp**: liczba członków rodziny pasażera znajdujących się na pokładzie
8. **Parch**: liczba członków rodziny pasażera znajdujących się na pokładzie
9. **Ticket**: sygnatura biletu pasażera,
10. **Fare**: cena biletu,
11. **Cabin**: numer zajmowanej kajuty,
12. **Embarked**: port załadunku pasażera

**Techniki analizy:**

1. **Random Forest Classifier**:
   * Do przewidywania przeżycia pasażerów zastosowano klasyfikator **Random Forest**. Jest to algorytm bazujący na zbiorze drzew decyzyjnych, który dzięki technice "bagging" (bootstrapping) pomaga w redukcji wariancji oraz poprawie dokładności predykcji. Algorytm Random Forest dobrze radzi sobie z danymi o różnych typach cech (kategorycznych i numerycznych) oraz może skutecznie radzić sobie z danymi brakującymi.
2. **LIME (Local Interpretable Model-Agnostic Explanations)**:
   * Algorytm LIME został zastosowany do wyjaśnienia decyzji modelu Random Forest. LIME generuje lokalne wyjaśnienia dla pojedynczych przypadków, wskazując, które cechy miały największy wpływ na przewidywanie dla danej próbki. Jest to przydatne w kontekście interpretacji modelu, który może być uznawany za "czarną skrzynkę", zwłaszcza w przypadku bardziej złożonych algorytmów, takich jak Random Forest.
3. **Ewaluacja modelu**:
   * **Raport klasyfikacji**: Raport ten zawiera miary wydajności modelu, takie jak dokładność (accuracy), precyzja (precision), recall (czułość) oraz F1-score. Dla każdej klasy (przeżył, nie przeżył) obliczane są te miary, a także średnia ważona.
   * **Macierz błędów**: Macierz błędów pokazuje liczbę prawidłowych oraz błędnych klasyfikacji (True Positives, True Negatives, False Positives, False Negatives). Jest to narzędzie pomagające ocenić, jak dobrze model radzi sobie z przewidywaniem różnych klas.
   * **Wizualizacja wpływu cech**: Zastosowanie wykresów słupkowych w celu przedstawienia wpływu poszczególnych cech na decyzje modelu. Cechy takie jak płeć, klasa podróży i wiek miały szczególnie duży wpływ na decyzje modelu.

# Wyniki i dyskusja

Po wytrenowaniu modelu klasyfikacyjnego Random Forest, przeprowadzona została jego ewaluacja na zbiorze testowym. Wyniki oceny modelu zostały przedstawione za pomocą raportu klasyfikacji oraz macierzy błędów, które stanowią podstawowe miary skuteczności modelu klasyfikacyjnego.

**Raport klasyfikacji:**

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

* **Precyzja**: Dla klasy "Not Survived" (0) wynosi 0.80, co oznacza, że 80% przypadków, które zostały przewidziane jako "nie przeżyli", zostały sklasyfikowane poprawnie. Dla klasy "Survived" (1) precyzja wynosi 0.75.
* **Recall (czułość)**: Dla klasy "Not Survived" recall wynosi 0.83, co oznacza, że model poprawnie przewidział 83% przypadków, które faktycznie nie przeżyły. Dla klasy "Survived" recall wynosi 0.71, co sugeruje, że model wykrył 71% przypadków przeżycia.
* **F1-score**: Dla klasy "Not Survived" wynosi 0.82, a dla klasy "Survived" wynosi 0.73, co pokazuje lepszą równowagę w przewidywaniu osób, które nie przeżyły, w porównaniu do tych, które przeżyły.

Dokładność modelu wyniosła 0.78, co oznacza, że model poprawnie sklasyfikował 78% przypadków w zbiorze testowym. Jest to miara ogólnej skuteczności modelu, określająca odsetek prawidłowych przewidywań w stosunku do wszystkich przypadków.

**Macierz błędów:**

Macierz błędów przedstawia liczbę prawidłowych oraz błędnych klasyfikacji.

Obraz zawierający tekst, zrzut ekranu, Oprogramowanie multimedialne, oprogramowanie

Opis wygenerowany automatycznie

Z tego wynika, że:

* TP = 79 (pasażerowie, którzy przeżyli i zostali poprawnie sklasyfikowani),
* TN = 130 (pasażerowie, którzy nie przeżyli i zostali poprawnie sklasyfikowani),
* FP = 27 (pasażerowie, którzy nie przeżyli, ale zostali błędnie sklasyfikowani jako "Survived"),
* FN = 32 (pasażerowie, którzy przeżyli, ale zostali błędnie sklasyfikowani jako "Not Survived")

Poniżej przedstawiono szczegółowe wyjaśnienie wyników klasyfikacji dla czterech różnych przypadków: **True Negative (TN)**, **False Positive (FP)**, **False Negative (FN)** i **True Positive (TP)**, uwzględniając prawdopodobieństwo oraz cechy mające największy wpływ na predykcję. Do każdego przypadku dołączono wykres feature influence, który ilustruje wpływ cech na decyzję modelu.

1. **True Negative (TN)**:
   * **True Label**: 0 (Not Survived)
   * **Predicted Label**: 0 (Not Survived)
   * **Observation Index**: 439
   * **Probabilities (Not Survived, Survived)**: [0.93, 0.07]

Model poprawnie przewidział, że pasażer nie przeżył katastrofy. Prawdopodobieństwo wynosiło [0.93, 0.07], co oznacza, że model był bardzo pewny swojej decyzji, przewidując 93% szans na "Not Survived".

**Cechy mające największy wpływ**:

* + **Płeć**: Mężczyzna, co miało duży wpływ na przewidywanie, ponieważ kobiety miały wyższe szanse na przeżycie.
  + **Klasa podróży**: Pasażer z klasą 3, która wiązała się z niższymi szansami przeżycia.

**Wizualizacja wpływu cech**:

Obraz zawierający zrzut ekranu, tekst, modelowanie 3D, Oprogramowanie graficzne

Opis wygenerowany automatycznie

1. **False Positive (FP)**:
   * **True Label**: 0 (Not Survived)
   * **Predicted Label**: 1 (Survived)
   * **Observation Index**: 396
   * **Probabilities (Not Survived, Survived)**: [0.45, 0.55]

W tym przypadku model błędnie przewidział, że pasażer przeżył. Prawdopodobieństwo wynosiło [0.45, 0.55], co oznacza, że model miał 55% szans na to, że pasażer przeżył.

**Cechy mające największy wpływ**:

* + **Płeć**: Kobieta, co znacząco wpłynęło na przewidywanie przeżycia.
  + **Wiek**: Młodszy pasażer również miał większe szanse na przeżycie, co mogło wpłynąć na decyzję modelu.

**Wizualizacja wpływu cech**:

Obraz zawierający zrzut ekranu, tekst, modelowanie 3D, Oprogramowanie graficzne

Opis wygenerowany automatycznie

1. **False Negative (FN)**:
   * **True Label**: 1 (Survived)
   * **Predicted Label**: 0 (Not Survived)
   * **Observation Index**: 709
   * **Probabilities (Not Survived, Survived)**: [0.71, 0.29]

Model błędnie przewidział, że pasażer nie przeżył, mimo że w rzeczywistości przeżył. Prawdopodobieństwo wynosiło [0.71, 0.29], co oznacza, że model przypisał 71% szans na to, że pasażer nie przeżył.

**Cechy mające największy wpływ**:

* + **Klasa podróży**: Pasażer z klasą 3, co obniżyło szanse przeżycia.

**Wizualizacja wpływu cech**:

Obraz zawierający zrzut ekranu, tekst, Oprogramowanie multimedialne, Oprogramowanie graficzne

Opis wygenerowany automatycznie

1. **True Positive (TP)**:
   * **True Label**: 1 (Survived)
   * **Predicted Label**: 1 (Survived)
   * **Observation Index**: 720
   * **Probabilities (Not Survived, Survived)**: [0.04, 0.96]

Model poprawnie przewidział, że pasażer przeżył katastrofę. Prawdopodobieństwo wynosiło [0.04, 0.96], co oznacza, że model był bardzo pewny swojej decyzji, przewidując 96% szans na "Survived".

**Cechy mające największy wpływ**:

* + **Płeć**: Kobieta, co miało duży wpływ na przewidywanie przeżycia.
  + **Klasa podróży**: Pasażer z klasą 1, co zwiększało szanse przeżycia.

**Wizualizacja wpływu cech**:

Obraz zawierający zrzut ekranu, tekst, modelowanie 3D, diagram

Opis wygenerowany automatycznie

# Podsumowanie

Celem projektu było zbudowanie modelu klasyfikacyjnego do przewidywania, czy pasażer Titanica przeżył katastrofę, na podstawie dostępnych cech, takich jak płeć, wiek, klasa podróży, liczba członków rodziny na pokładzie oraz inne informacje dotyczące pasażerów. Do realizacji zadania wykorzystano klasyfikator **Random Forest**, który skutecznie przewidywał przeżycie pasażerów na podstawie danych.

Model osiągnął dokładność **78%**, co wskazuje na jego stosunkowo dobrą skuteczność, ale jednocześnie wskazuje na pewne trudności w przewidywaniu przeżycia pasażerów, którzy faktycznie przeżyli katastrofę. W szczególności, model miał większą liczbę False Negatives (32), co oznacza, że pomimo poprawnego rozpoznania wielu pasażerów, którzy nie przeżyli, nie udało się przewidzieć przeżycia części pasażerów, którzy w rzeczywistości przeżyli.

Zastosowanie **algorytmu LIME** pozwoliło na głębsze zrozumienie, jakie cechy miały największy wpływ na decyzje modelu. Wśród najistotniejszych cech wymieniono płeć i klasę podróży. Kobiety oraz pasażerowie klasy 1 mieli większe szanse na przeżycie, co miało wyraźny wpływ na przewidywania modelu. Cechy te były szczególnie istotne w przypadkach poprawnie sklasyfikowanych jako "Survived" (True Positive), ale także w przypadkach błędnie klasyfikowanych jako "Not Survived" (False Negative).

Wizualizacje wpływu cech, wygenerowane przy użyciu algorytmu LIME, pozwoliły na lepsze zrozumienie, które zmienne miały dominujący wpływ na wyniki modelu. Pomogło to w ocenie, czy model podejmuje decyzje zgodnie z naszymi oczekiwaniami, zwłaszcza w kontekście przewidywania przeżycia pasażerów.

# Bibliografia

https://www.kaggle.com/code/prashant111/explain-your-model-predictions-with-lime

youtube: Applying LIME with Python | Local & Global Interpretations