

Projekt – Systemy Sztucznej Inteligencji

Zadanie 1

Na podstawie zbioru danych Travel Insurance Prediction wykonaj poniższe zadania:

1. Policz średnią arytmetyczną wieku klientów
2. Policz medianę przychodu klientów, którzy mają status Frequent Flyer.
3. Policz ilu klientów ma choroby przewlekłe
4. Usuń kolumnę "Id" ze zbioru
5. Do pliku zapisz następujące dane klientów: wiek, przychód, liczba członków rodziny, choroby przewlekłe.
6. Wybierz klientów, którzy kupili ubezpieczenie i zapisz dane klientów do pliku (bez kolumny TravelInsurance)
7. Wybierz klientów, których przychód jest większy niż średnia
8. Stwórz nową kolumnę zawierającą dane kategoryjne: "less 30", "30-39", "40-49", "50+" na podstawie kolumny zawierającej wiek

Zadanie 2

Zaimplementuj następujące transformery dla danych ze zbioru Travel Insurance Prediction:

1. Transformer wybierający informacje o klientach, których dochód jest większy niż średnia arytmetyczna
2. Transformer zamieniający wartość Yes/No na wartości 0/1 we wszystkich kolumnach, w których występują.
3. Transformer usuwający podane kolumny

Stwórz pipeline, w którym przetestujesz powyższe transformery

Projekt

Celem projektu jest stworzenie aplikacji, która będzie wykorzystywała model sztucznej inteligencji w problemie klasyfikacji (binarnej lub wieloklasowej). Aplikacja nie musi posiadać GUI, może być konsolowa.

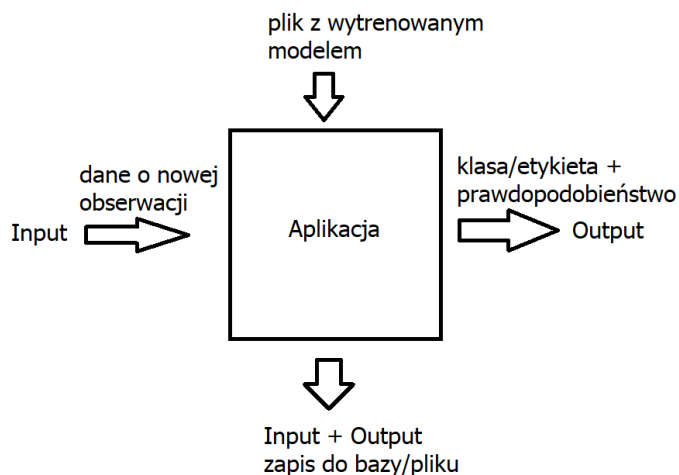
Przykład: aplikacja do przewidywania czy klient dostanie kredyt. Do aplikacji wprowadzane są dane klienta przez użytkownika, a jako odpowiedź aplikacja zwraca informację czy klient otrzyma kredyt i z jakim prawdopodobieństwem.

Przed dołączeniem modelu do aplikacji należy przetestować model.

Testowanie modelu: podział na zbiór treningowy i testowy, przetestowanie kilku różnych klasyfikatorów, obliczenie i porównanie metryk, przetestowanie modelu w oparciu o cross walidację.

Następnie należy wybrać jeden klasyfikator i wytrenować model na całym zbiorze danych, zapisując wytrenowany model do pliku. Proces testowania modelu ma zostać zrealizowany w oddzielnym notatniku. Wytrenowany model ma zostać zapisany do pliku.

Dane, które zostały wprowadzone do aplikacji wraz z odpowiedzią modelu i z prawdopodobieństwem tej odpowiedzi powinny zostać zapisane do pliku/bazy danych.



Projekt będzie prezentowany na ostatnich zajęciach. Projekt należy zrobić w grupach dwuosobowych.

Zadanie 3

Znajdź zbiór danych dla swojego projektu. Zbiór danych powinien zawierać dane jakościowe i ilościowe. Zbiór danych ma być przystosowany do problemu klasyfikacji.

Przedstaw krótki opis zbioru danych: ile cech zawiera zbiór, jakiego rodzaju są cechy: ilościowe czy jakościowe, ile obserwacji liczy zbiór, dla wartości jakościowych przedstaw unikalne wartości, dla wartości ilościowych przedstaw podstawowe statystyki. Podaj cechę zależną i cechy niezależne. Proces eksploracji danych (w tym opis) ma zostać zrealizowany w oddzielnym notatniku.

Zadanie 4

Przygotuj zbiór danych do uczenia maszynowego, m.in:

- Usuń niepotrzebne kolumny,
- Zmień nazwy kolumn jeśli są np. nieczytelne,
- Zakoduj zmienne kategoryjne za pomocą one-hot,
- Dokonaj przekształceń innych przekształceń jeśli są potrzebne,
- Zapisz zbiór danych do pliku csv lub xlsx.

Na potrzeby projektu obserwacje z wartościami pustymi należy usunąć.

Zadanie 5

Zaprojektuj aplikację oraz format pliku/bazy danych, tak aby w następnym kroku można było podpiąć wytrenowany model.

Zadanie 6

Wytrenuj model, który podłączysz do aplikacji. Przed ostatecznym wytrenowaniem modelu, przetestuj różne klasyfikatory, zastosuj cross walidację, oblicz i porównaj różne metryki. Stwórz oddzielny notatnik, w którym przetestujesz różne modele i wytrenujesz ostateczny model. Notatnik ma zawierać opis co się w nim znajduje.

Zadanie 7

Podepnij wytrenowany model do aplikacji. Dodaj zapisywanie do pliku/bazy. Przetestuj działanie aplikacji.

Ocenianie:

- Ocena 3 – brak aplikacji, znalezienie zbioru danych, notatnik z eksploracją danych, notatnik z przygotowaniem danych do trenowania, notatnik z trenowaniem i testowaniem modelu,
- Ocena 4 – aplikacja konsolowa, znalezienie zbioru danych, notatnik z eksploracją danych, notatnik z przygotowaniem danych do trenowania, notatnik z trenowaniem i testowaniem modelu, zapisywanie danych do pliku,
- Ocena 5 – aplikacja z GUI, znalezienie zbioru danych, notatnik z eksploracją danych, notatnik z przygotowaniem danych do trenowania, notatnik z trenowaniem i testowaniem modelu, zapisywanie danych do bazy.

Co będzie podlegać ocenie?

- Czytelność i jakość kodu,
- Dyskusja przy prezentowaniu projektu.