|  |  |  |  |
| --- | --- | --- | --- |
|  | Politechnika Bydgoska im. J. J. Śniadeckich  Wydział Telekomunikacji,  Informatyki i Elektrotechniki  **Zakład Systemów Teleinformatycznych** | |  |
| **Przedmiot** | Zaawansowane Techniki Sztucznej Inteligencji | | |
| **Prowadzący** | prof. dr hab. inż. prof. PBŚ Piotr Cofta | | |
| **Temat** | *Project* | | |
| **Student** | Cezary Tytko | | |
| **Ocena** |  | **Data oddania spr.** |  |

Etap 1. – kontrola danych.

Weryfikacji możemy dokonać przez wczytanie danych jako pliku csv i sprawdzić czy wymiary otrzymanej tablicy zgadzają się z przewidywanymi, to znaczy czy mamy odpowiednia liczbę rekordów i czy każdy rekord zawiera odpowiednią liczbę danych, to jest suma liczby oczek na widocznych poprawnych kostkach i 100 X 100 pikseli ułożonych w jednym wymiarze. Należy również sprawdzić czy wartości etykiet jak o danych są zgodne z założeniami, np. czy nie wychodzą poza przewidziany zakres, albo czy nie ma wartości brakujących, można to sprawdzić wyświetlając wartości unikatowe (Select distinct w konwencji sql), zgodnie z założeniami wartość etykiety powinna być z zakresu <6, 30>, a dla piksela <0, 255>, jeżeli dane będą zawierały błędy na poziome typów wartości tzn. string nie konwertowany na int, to dostaniemy błąd na etapie odczytu pliku csv (przynajmniej dla implementacji z pandas).

Przykładowy kod weryfikacji:

1. dice\_y, dice\_x = ReadDiceCSV(dice5\_dir)

2. dice\_x = dice\_x.reshape((dice5\_count, 100, -1))

3. print("dice:")

4. print(f"Label Shape: {dice\_y.shape} has NAN: {np.isnan(dice\_y).any()}")

5. print(f"Values : {np.unique(dice\_y)}")

6. print(f"Data Shape: {dice\_x.shape} has NAN: {np.isnan(dice\_x).any()}")

7. print(f"Values : {np.unique(dice\_x)}")

8.

Wynik weryfikacji:

Obraz zawierający tekst, zrzut ekranu, Czcionka, czarne i białe

Opis wygenerowany automatycznie

Weryfikacji pokazała nie zgodność rzeczywistych danych z wyżej wymienionymi założeniami jakie powinny spełniać, ponieważ rzeczywiste etykiety są z zakresu <5, 30>, a założenie było <6, 30>, wartości pikseli są zgodne z założeniami.

Przykładowy wyświetlony obraz zgodny z założeniem (nr = index modulo liczba obrazów):

Obraz zawierający tekst, wzór, Prostokąt, zrzut ekranu

Opis wygenerowany automatycznie

Sprawdziłem dodatkowo liczność wystąpień i utworzyłem z nich histogram:

Obraz zawierający zrzut ekranu, tekst, diagram, Wykres

Opis wygenerowany automatycznie

Jak się spodziewałem większość wartości jest średnią z minimum i maksimum (5 + 30) / 2 = 17.5, zatem najliczniejsze klasy to suma wynosząca 17 i 18 oczek.