

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



Klasyfikacja oparta na danych dotyczących ocen gier

Sprawozdanie z laboratorium

AUTOR

Stsiapan Shkliar

nr albumu: **260467**

kierunek: **Informatyka Stosowana**

13 czerwca 2022

Streszczenie

Celem projektu jest predykcja na podstawie danych wcześniejszych ocen czy danemu użytkownikowi spodoba się dana gra czy nie. Dane były pobrane <https://howlongtobeat.com> oraz <https://www.metacritic.com>. Pobrane dane zostały następnie oczyszczone do predykcji oceny gry na podstawie jej cech.

1 Wstęp – sformułowanie problemu

Jaki jest problem? Autor potrzebuje przewidzieć, czy spodoba się mu ta gra, czy nie. Pozwoli mu to na predykowanie oceny gry oraz czy warto taką grę kupować czy nie...

2 Opis danych

Wielkość datasetu 300 wierszy. Kolumna "Title" - zmienna typu string, określa ona nazwę gry. Zbiór wartości: Lumines, Puzzle Fusion, Mr. DRILLER: Drill Spirits, Call of Duty 2, Need for Speed: Most Wanted 5-1-0.

Kolumna "Max Players" - zmienna typu 'int', określa ona maksymalną liczbę graczy. Zbiór wartości: 1, 2, 4.

Kolumna "Multiplatform" - zmienna typu 'boolean', określa ona czy dana gra potrzymuje multiplatform. Zbiór wartości: true, false.

Kolumna "Online" - zmienna typu 'boolean', określa ona czy można grać online w danej grze. Zbiór wartości: true, false.

Kolumna "Genres" - zmienna kategoryczna, określa ona gatunek gry. Zbiór wartości: Action, Strategy, Sports, Adventure, RPG.

Kolumna "Publishers" - zmienna kategoryczna, określa ona wydawców gry. Zbiór wartości: Nintendo, Ubisoft, EA, Adventure, Konami.

Kolumna "Review Score" - zmienna typu 'int', określa ona ocenę gry. Zbiór wartości: 85, 89, 35, 68.

Kolumna "Total Sales" - zmienna typu 'float', określa ona sumę z sprzedaży gry graczy. Zbiór wartości: 4.69, 0.56, 0.41, 0.06.

Kolumna "Price" - zmienna typu 'int', określa ona maksymalną liczbę graczy. Zbiór wartości: 24.95, 29.99, 41.99.

Kolumna "ESRB Rating" - zmienna kategoryczna, określa ona ocenę ESRB. Zbiór wartości: E, M, A, RP.

Kolumna "Year of Release" - zmienna typu 'int', określa ona rok wydania gry. Zbiór wartości: 2001, 2015, 2008.

Kolumna "Number Of Players Completed The Game" - zmienna typu 'int', określa ona liczbę graczy, którzy skończyli grę. Zbiór wartości: 57, 888, 1020.

Kolumna "Completionists Average Time" - zmienna typu 'float', określa ona średnią liczbę godzin zużytych przez gracza dla skończenia gryw całości. Zbiór wartości: 29.766666666666666, 1.25, 16.383333333333333.

Kolumna "Main Story Average Time" - zmienna typu 'int', określa ona średnią liczbę godzin zużytych przez gracza dla skończenia głównie części gry. Zbiór wartości: 14.333333333333334, 10.333333333333334, 15.5.

Kolumna "Like" - zmienna typu 'int', określa ona czy podoba się dana gra. Zbiór wartości: 0, 1.

3 Opis rozwiązania

Dane zostały stworzone na podstawie znalezionej informacji ze stron <https://howlongtobeat.com>, <https://www.metacritic.com> oraz opinii użytkownika. Baza została zapisana w postaci ramki

danych biblioteki **Pandas**. Zawiera ona informacje o 300 gier wraz z 14 cechami je określającymi.

Używając metody klasyfikacji za pomocą *Neural Neu* na danych uzyskano model pozwalający na określenie oceny gry na podstawie jego cech.

4 Rezultaty obliczeń

4.1 Plan badań

Zbiór danych zostanie podzielony na dwie części: treningową i testową w stosunku 80:20.

4.2 Metryki używane przy klasyfikacji

Dokładność (Accuracy)

Metryka poprawności zaklasyfikowanych elementów we wszystkich klasach.

Dokładność (Accuracy)

Metryka poprawności zaklasyfikowanych elementów we wszystkich klasach.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precyzja (Precision/PPV)

Metryka jakości zaklasyfikowanych obiektów (Jak relewantne są wyniki).

$$Precision = \frac{TP}{TP + FP}$$

Czułość (Recall)

Metryka wskazująca ile obiektów danej klasy zostało poprawnie zaklasyfikowane.

$$Recall = \frac{TP}{TP + FN}$$

F1 score

F1-score: średnia harmoniczna z precyzji i czułości: Miara ta daje ocenę balansu między czułością a precyzją. Miara ta nie uwzględnia wyników prawdziwie negatywnych.

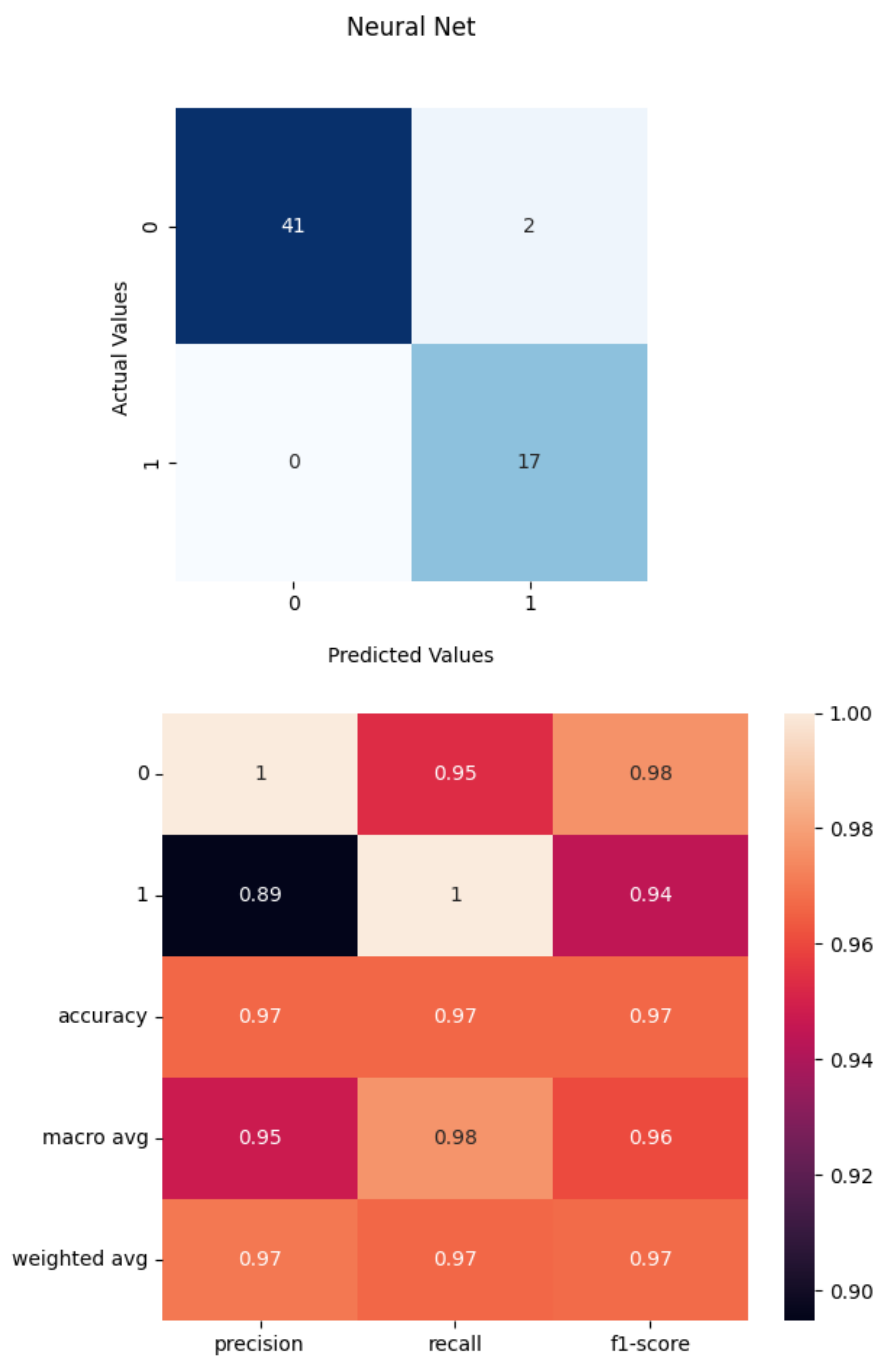
$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Współczynnik korelacji Matthews (Matthews correlation coefficient)

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Zbalansowana miara korelacji pozwalająca zbadać jakość klasyfikacji w niezbalansowanym datasetcie.

4.3 Wykresy obliczeń



5 Wnioski

Przedstawiona klasyfikacja działa poprawnie z dobrym procentem dokładności na podstawie danych wcześniejszych ocen, czyli może znaleźć grę, która spodoba się użytkownikowi. W przyszłości klasyfikacja ta może zostać rozszerzona na dobry system rekomendacji.

A Dodatek

Kody źródłowe (utrzymane w konwencji języka Python wraz z instrukcjami uruchomienia) umieszczone zostały w repozytorium github:
https://github.com/nonamestephan/MSID_project.