

YouTube trending videos

Autorzy: **Bartosz Górka** INF127228 bartosz.gorka@student.put.poznan.pl
Kajetan Zimniak INF127229 kajetan.zimniak@student.put.poznan.pl

22 marca 2020

1 Wstęp

Celem projektu było przeprowadzenie procesu odkrywania wiedzy z rzeczywistych danych. W szczególności: pozyskania danych z różnych źródeł, przetworzenia ich do dogodnej reprezentacji, oceny jakości danych, oceny ważności atrybutów, poszukiwanie współzależności między nimi, odkrycia użytecznych i potencjalnie interesujących regularności, oraz dokonanie interpretacji znalezionych regularności. Szczegółowy opis zadania znajduje się na stronie mgr Konrada Miazgi[1]. Szczegółowe analizy znajdują się w pliku Jupyter Notebook.

2 Atrybuty tekstowe

Pierwszy etap projektu polega na przygotowaniu wstępnych statystyk danych, wykorzystaniu metod wizualizacji, zapoznaniu się z danymi oraz ich jakością. Bierze się pod uwagę atrybuty tekstowe i liczbowe, z pominięciem obrazków.

W pierwszej kolejności wprowadzono dodatkowe atrybuty, które miały służyć pewnego rodzaju normalizacji atrybutów liczbowych. Dodano stosunek liczby polubień, łapek w dół oraz komentarzy do liczby wyświetleń, oraz stosunek liczby polubień do łapek w dół.

Przedstawiono również statystyki dotyczące liczby wyświetleń w zależności od pory dnia i dnia tygodnia, sprawdzono czy istnieje różnica pomiędzy USA a Wielką Brytanią, oraz zbadano tendencję aktywności użytkowników na przestrzeni badanych lat.

Podczas analizy tytułów i tagów zwrócono szczególnie uwagę na ich długość (zarówno liczbę słów jak i znaków), powtórzenia, liczbę użytych punktorów ('!', '?', '...' itd.), wielkość liter, częstotliwość występowania tzw. *stopwords*. Uznano, że ważnym elementem jest również dobór słów używanych w tytułach, tagach i opisach. Sporządzono ranking najczęstszych z nich.

Sprawdzono także, które kategorie cieszą się największą popularnością i ile filmów zezwala na komentarze i oceny.

Analizie poddano również korelację pomiędzy atrybutami.

3 Atrybuty wizualne

Literatura

[1] mgr Konrad Miazga. URL <https://www.cs.put.poznan.pl/kmiazga/ped.html>.