Break the FAKE

**Programisto eventu HackYeah!**

Wyobraź sobie, że codziennie na skrzynkę e-mail dostajesz gazetę, ale nie taką samą jak wszyscy. Ta została napisana specjalnie dla Ciebie i zawiera tylko prawdziwe, zweryfikowane i aktualne informacje gospodarcze i ekonomiczne z wielu źródeł, którym możesz zaufać. Chcesz odcisnąć ślad na polskiej rzeczywistości? Pomóż nam dostarczać wiarygodne informacje i zatrzymaj fake newsy. Podejmij wyzwanie i **Break the Fake!**

**O naszym zadaniu**

Rosnąca ilość informacji generuje coraz to większy szum utrudniający wyłapywanie fake newsów. W tej edycji oczekujemy od Was zaprojektowania rozwiązania, które będzie w stanie ustrukturyzować codzienne newsy, a także ocenić ich sentyment oraz określić prawdopodobieństwo, że dana informacja jest nieprawdziwa. To zadanie sprawdzi praktyczne umiejętności programowania, a duża swoboda w jego realizacji nie będzie ograniczać Waszej kreatywności.

Oczekujemy zaprojektowania rozwiązania, bazującego na językach Python i/lub R, które będzie potrafił obsługiwać każdy użytkownik. Ma ono umożliwiać przeglądanie artykułów napisanych w języku polskim i pochodzących z wielu źródeł. **Rozwiązanie musi klasyfikować każdy artykuł na podstawie jego treści do co najmniej jednej kategorii przedstawionych załączniku (1).**

Warunki jakie musi spełnić projektowane przez Was rozwiązanie zostały określone w dwóch kategoriach – konieczne i dodatkowe.

**Nasze wymagania**

I Aplikacja powinna(warunki **konieczne**):

1. Posiadać możliwie jak najprostszy, przyjazny, funkcjonalny i jak najbardziej intuicyjny interfejs, umożliwiający filtrowanie według dat, kategorii, źródła, oceny sentymentu, prawdopodobieństwa *fake newsa*.
2. Być dostosowana do zastosowania na systemie operacyjnym Windows 10.
3. Wyświetlać pozycje artykułów wraz z następującymi informacjami:
   1. tytuł,
   2. streszczenie,
   3. źródło artykułu (nazwa źródła oraz link),
   4. czas publikacji,
   5. autora artykułu,
   6. tagi,
   7. kategorię (dokonaną przez algorytm klasyfikacyjny),
   8. określenie sentymentu dotyczący przekazu (pozytywny, neutralny, negatywny),
   9. ocenę prawdopodobieństwa, że artykuł jest *fake newsem* (maksymalnie do 5 kategorii).
4. archiwizować zebrane artykuły.
5. umożliwiać ustawienie częstości zasilania archiwum artykułów.

Dostarczone rozwiązanie musi bazować na narzędziach, których dalsze wykorzystanie przez Ministerstwo Finansów nie będzie wiązało się z ponoszeniem dodatkowych kosztów.

II Aplikacja może zapewniać dodatkowe funkcje, które w istotny sposób wpływają na funkcjonalność interfejsu (warunki **dodatkowe**) np.

* możliwość tworzenia wykresów dla poszczególnych możliwości filtrowania,
* określenie pierwotnego pochodzenia artykułu, w tym grafowa wizualizacja,
* określenie sentymentu w sposób zaawansowany np. emocje,
* ocena prawdopodobieństwa, że w artykule znajduje się „mowa nienawiści”.

**Ocena Waszych projektów**

Oceniany jest przygotowany interfejs, który uczestnik Konkursu lub Zespół zaprezentuje przed komisją. Do oceny prezentowanego rozwiązania interfejs ma przedstawiać artykuły z maksimum ostatnich 30 dni. Ocena składa się z dwóch etapów.

**Etap 1**

Oceniane jest spełnienie warunków koniecznych, a następnie podejmowana jest decyzja o zakwalifikowaniu rozwiązania do etapu II oceny.

**Etap 2**

Oceniana jest jakość dostarczonego rozwiązania wraz z dodatkowymi funkcjonalnościami na skali punktowej od 0 do 100 punktów, na które składają się:

1. stopień realizacji i jakość poszczególnych składowych funkcjonalności rozwiązania (0-40). Kryteria oceny będą uwzględniały w szczególności elementy: liczba źródeł danych, liczba artykułów, liczba i trafność klasyfikacji artykułu, liczba i trafność określenia sentymentu artykułów, liczba i trafność oceny prawdopodobieństwa artykułu jako *fake newsa*,
2. intuicyjność obsługi zaprojektowanego interfejsu i jego ergonomia (0-20). Kryteria oceny będą uwzględniały przyjazność obsługi dla użytkownika o podstawowych umiejętnościach korzystania z komputera oraz estetykę interfejsu,
3. architektura rozwiązania oraz możliwość wdrożenia systemu w środowisku produkcyjnym: punktacja (0 do 20). Punktacja będzie uwzględniała ocenę takich elementów jak: kompozycja modułów systemu, skalowalność, elastyczność względem różnych źródeł danych, zagadnienia dotyczące monitoringu pracy aplikacji, wydajności, rozliczalności, przenośności kodu,
4. poziom złożoności wykorzystanej metody klasyfikacji sentymentu oraz oceny artykułu jako *fake newsa* (0-10). Kryteria oceny będą uwzględniały adekwatność zastosowanej metody klasyfikacji oraz jej zaawansowania,
5. zapewnienie dodatkowych funkcji niewskazanych w wymaganiach, które   
   w istotny sposób wpływają na funkcjonalność interfejsu (0-10), np. możliwość tworzenia wykresów podsumowujących dla poszczególnych kategorii filtrów, określenie pierwotnego pochodzenia artykułu (w tym grafowa wizualizacja), określenie sentymentu w sposób zaawansowany np. emocje.

**Informacje dodatkowe**

Wszystkie pliki dotyczące rozwiązania muszą zostać dostarczone w formie pliku .zip o nazwie „Break-the-FAKE-*numerzespołu*” i zawierać co najmniej poniższe elementy:

1. plik o nazwie „INFO.zip” – archiwum zawierające wszelkie informacje dotyczące dostarczonego rozwiązania oraz ewentualnej konfiguracji środowiska, a także opis kroków umożliwiających uruchomienie interfejsu,
2. plik o nazwie „CODE.zip” – archiwum zawierające kody źródłowe lub adresy lokalizacji plików z kodami źródłowymi pozwalające na uzyskanie skompilowanej, działającej wersji oprogramowania,
3. plik o nazwie „DASHBOARD.zip” – archiwum zawierające skompilowaną, na podstawie kodów źródłowych zawartych w pliku „CODE.ZIP”, wersję narzędzia lub odnośnik z jego lokalizacją,
4. plik o nazwie „OTHERS.zip” – archiwum zawierające wszelkie inne informacje.