

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ ELEKTRONIKI

KIERUNEK: INFORMATYKA
SPECJALNOŚĆ: Inżynieria systemów informatycznych

PRACA DYPLOMOWA
MAGISTERSKA

Zastosowanie metod uczenia maszynowego w
detekcji fałszywych informacji

Application of machine learning methods to
fake news detection

AUTOR:

Inż. Dawid Mikowski

PROWADZĄCY PRACĘ:

Prof. Michał Woźniak

OCENA PRACY:

Spis treści

Spis rysunków	4
Spis tabel	5
1. Streszczenie	6
2. Wstęp	7
2.1. Motywacja	7
3. Informacje nieprawdziwe w dobie internetu	8
3.1. Sposoby rozprzestrzeniania fałszywych informacji	9
3.1.1. Gazety	9
3.1.2. Telewizja	10
3.1.3. Internet	10
3.2. Deep fake	10
3.3. Sposoby ochrony przed nieprawdziwymi informacjami	10
4. Uczenie maszyn	11
4.1. Rodzaje	11
4.2. Algorytmy klasyfikacji	11
4.2.1. KNN	11
4.2.2. SVC	11
4.2.3. MLP	11
4.2.4. Binary trees	11
4.2.5. Naive Bayes	11
4.3. Wykorzystanie	11
4.4. Zagrożenia	11
5. Przetwarzanie języków naturalnych	12
5.1. Normalizacja danych tekstowych	12
5.2. Wektoryzacja	12
5.2.1. Bag of words	12
5.2.2. TfIDF	12
6. Projekt i implementacja systemu	13
6.1. Wykorzystane technologie	13
6.2. Wymagania funkcjonalne	13
6.3. Implementacja	13
7. Ocena eksperymentalna	14
7.1. Cel Badań	14

7.2. Warunki przeprowadzonego eksperymentu	14
7.3. Wyniki	14
7.4. Analiza wyników wraz z oceną statystyczną	14
7.5. Wnioski z badań	14
8. Podsumowanie	15
Literatura	16
A. Opis załączonej płyty CD/DVD	17

Spis rysunków

Rys. 3.1. Post udostępniony przez Donalda Trumpa na portalu Twitter	8
Rys. 3.2. Przykład żółtej prasy z roku 1993	10

Spis tabel

Rozdział 1

Streszczenie

Rozdział 2

Wstęp

2.1. Motywacja

Rozdział 3

Informacje nieprawdziwe w dobie internetu

Pojęcie “Fake news” odnosi się do informacji, które pomimo że nie posiadają pokrycia z rzeczywistością są przedstawiane jako prawdziwe w mediach takich jak np.: wiadomości, artykuły, portale społecznościowe itd.. Zwrot ten jest neologizmem i w języku angielskim oznacza dosłownie “Fałszywe wiadomości” celem wykorzystania takich wiadomości mogą być żarty np.: satyra, jednak najczęściej mają one jedno z dwóch zadań:

- oszukać odbiorcę i wpłynąć na jego poglądy w sposób żądany przez autora danej informacji *Propaganda*
- namówienie go na zakup czegoś czego w innym przypadku by on nie kupił *Reklama*

Niektóre fałszywe informacje łączą oba cele.

Fake news’y stały się bardzo popularnym zagadnieniem w ostatnich czasach ponieważ internet a w szczególności media społecznościowe pozwoliły na przekazywanie informacji z niespotykaną wcześniej prędkością dzięki czemu rozprzestrzenianie dezinformacji stało się zadaniem stosunkowo prostym.

Zagadnienie to zyskało ogromny rozgłos podczas kampanii wyborczej oraz prezydentury Donalda Trumpa, który zasłynął z częstego wykorzystywania tego zwrotu podczas wywiadów, debat oraz wypowiedzi na mediach społecznościowych tj. Twitter. Do roku 2020 pojęcie “Fake News” zostało umieszczone w słownikach języka angielskiego takich jak “Oxford English Dictionary”, “Macmillan Dictionary”.



Rys. 3.1: Post udostępniony przez Donalda Trumpa na portalu Twitter

Według założonego przez dziewięć organizacji w skład których wchodzi Google, Facebook oraz Twitter projektu “First Draft News” możemy wyróżnić siedem typów Fake Newsów:

1. Satyra bądź parodia
2. Fałszywe połączenie
3. Myląca zawartość
4. Fałszywy kontekst
5. Oszukana zawartość
6. Zmanipulowana zawartość
7. Sfabrykowana zawartość

Jak podaje słownik “Merriam Webster” po raz pierwszy wykorzystano zwrot “Fake news” w roku 1890.

Jednym z najsłynniejszych typów fake newsów jest propaganda, czyli według definicji “technika sterowania poglądami i zachowaniami ludzi polegająca na celowym, natarczywym, połączonym z manipulacją oddziaływaniu na zbiorowość” pomimo iż najczęściej propaganda ma charakter polityczny nie jest to jedyne jej zastosowanie. Najstarszym przykładem pisemnej propagandy są opisy podbojów Dariusza Wielkiego datowane na rok 515 p.n.e. Od tego czasu w historii ludzkości można znaleźć wiele przypadków wykorzystania tego typu dezinformacji w takich krajach jak Starożytny Rzym, Niemcy podczas drugiej wojny światowej a nawet w dzisiejszych czasach Korea Północna. Propagandę można podzielić na 3 różne typy:

1. Biała propaganda - źródło pochodzenia informacji jest prawdziwe i podane
2. Szara propaganda - źródło pochodzenia informacji jest dla odbiorcy nieznane i może się on jedynie domyślać
3. Czarna propaganda - źródło pochodzenia informacji jest umyślnie sfałszowane w celu wyrządzenia szkody

3.1. Sposoby rozprzestrzeniania fałszywych informacji

Wraz ze zmianami w sposobach rozprzestrzeniania informacji na świecie zmieniało się także podejście do tworzenia fake newsów w odpowiedni sposób oszukujących osoby do których były one skierowane.

3.1.1. Gazety

Wykorzystanie fake newsów w gazetach miało głównie na celu przyciągnąć uwagę a co za tym idzie zwiększyć sprzedaż danej gazety. Stało się to na tyle popularne że spowodowało narodziny nowego pojęcia “żółtej prasy” była to prasa starająca się z całych sił zwrócić uwagę przechodnia poświęcając swoją wiarygodność poprzez zawarcie w nagłówkach w pełni lub częściowo nieprawdziwych wiadomości. Dziennikarz Frank Luther Mott wyróżnia 5 cech charakteryzujących żółtą prasę:

- Napisane dużą czcionką straszące nagłówki na temat mniej ważnych wydarzeń
- Nadmierna ilość zdjęć i rysunków
- Zawarcie sfałszowanych wywiadów, mylących nagłówków, pseudonauki oraz nieprawdziwych informacji od ludzi podających się za ekspertów
- Dodanie w pełni kolorowych dodatków do gazet w niedzielę
- Stawianie siebie jako słabszego w walce przeciwko systemowi



Rys. 3.2: Przykład żółtej prasy z roku 1993

3.1.2. Telewizja

3.1.3. Internet

3.2. Deep fake

3.3. Sposoby ochrony przed nieprawdziwymi informacjami

Rozdział 4

Uczenie maszyn

4.1. Rodzaje

4.2. Algorytmy klasyfikacji

4.2.1. KNN

4.2.2. SVC

4.2.3. MLP

4.2.4. Binary trees

4.2.5. Naive Bayes

4.3. Wykorzystanie

4.4. Zagrożenia

Rozdział 5

Przetwarzanie języków naturalnych

5.1. Normalizacja danych tekstowych

5.2. Wektoryzacja

5.2.1. Bag of words

5.2.2. TfIDF

Rozdział 6

Projekt i implementacja systemu

6.1. Wykorzystane technologie

6.2. Wymagania funkcjonalne

6.3. Implementacja

Rozdział 7

Ocena eksperymentalna

7.1. Cel Badań

7.2. Warunki przeprowadzonego eksperymentu

7.3. Wyniki

7.4. Analiza wyników wraz z oceną statystyczną

7.5. Wnioski z badań

Rozdział 8

Podsumowanie

Literatura

Dodatek A

Opis załączonej płyty CD/DVD

Na załączonej płycie znajduje się niniejsza praca w formacie PDF oraz pliki z kodem źródłowym aplikacji wykorzystanej do wykonania badań.