

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

<b>Spis rysunków</b>	<b>4</b>
<b>Spis tabel</b>	<b>5</b>
<b>1. Streszczenie</b>	<b>6</b>
<b>2. Wstęp</b>	<b>7</b>
2.1. Motywacja	7
<b>3. Informacje nieprawdziwe w dobie internetu</b>	<b>8</b>
3.1. Sposoby rozprzestrzeniania fałszywych informacji	9
3.1.1. Gazety	9
3.1.2. Telewizja	10
3.1.3. Internet	10
3.2. Sposoby ochrony przed nieprawdziwymi informacjami	12
3.3. Deep fake	13
<b>4. Uczenie maszyn</b>	<b>14</b>
4.1. Rodzaje	14
4.2. Algorytmy klasyfikacji	14
4.2.1. KNN	14
4.2.2. SVC	14
4.2.3. MLP	14
4.2.4. Binary trees	14
4.2.5. Naive Bayes	14
4.3. Wykorzystanie	14
4.4. Zagrożenia	14
<b>5. Przetwarzanie języków naturalnych</b>	<b>15</b>
5.1. Normalizacja danych tekstowych	15
5.2. Wektoryzacja	15
5.2.1. Bag of words	15
5.2.2. TfIDF	15
<b>6. Projekt i implementacja systemu</b>	<b>16</b>
6.1. Wykorzystane technologie	16
6.2. Wymagania funkcjonalne	16
6.3. Implementacja	16
<b>7. Ocena eksperymentalna</b>	<b>17</b>
7.1. Cel Badań	17

---

7.2. Warunki przeprowadzonego eksperymentu . . . . .	17
7.3. Wyniki . . . . .	17
7.4. Analiza wyników wraz z oceną statystyczną . . . . .	17
7.5. Wnioski z badań . . . . .	17
<b>8. Podsumowanie . . . . .</b>	<b>18</b>
<b>Literatura . . . . .</b>	<b>19</b>
<b>A. Opis załączonej płyty CD/DVD . . . . .</b>	<b>20</b>

# Spis rysunków

Rys. 3.1. Post udostępniony przez Donalda Trumpa na portalu Twitter . . . . .	8
Rys. 3.2. Przykład żółtej prasy z roku 1993 . . . . .	10
Rys. 3.3. Przykład fałszywej informacji na portalu facebook . . . . .	11
Rys. 3.4. Infografika stworzona przez IFLA . . . . .	12

# 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]:



1. Satyra bądź parodia - nie ma na celu wyrządzić krzywdy ale może oszukać
2. Fałszywe połączenie - nagłówki oraz obrazy nie mają powiązania z zawartością
3. Myląca zawartość - tekst napisany w mylący sposób
4. Fałszywy kontekst - prawdziwa zawartość powiązana ze złym kontekstem
5. Oszukana zawartość - źródła pochodzenia informacji są fałszywe
6. Zmanipulowana zawartość - prawdziwa zawartość zmanipulowana w odpowiedni sposób by oszukać odbiorcę
7. Sfabrykowana zawartość - całkowicie zmyślona zawartość mająca na celu wyrządzić krzywdę

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” [4] 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 sfalszowane 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 [7]:

- Napisane dużą czcionką straszące nagłówki na temat mniej ważnych wydarzeń
- Nadmierna ilość zdjęć i rysunków
- Zawarcie sfalszowanych wywiadów, mylących nagłówek, 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

Wynalezienie telewizji na początku 20 wieku zmieniło całkowicie sposób w jaki ludzie pozyskiwali wiadomości ze świata. Aby pozyskać informacje na temat najnowszych wydarzeń nie było potrzebne kupienie gazety a nawet wyjście z domu, w tym celu wystarczyło posiadanie telewizji i jej uruchomienie. Połączenie zarówno obrazu jak i dźwięku zmusiło osoby chcące oszukać swoich odbiorców do stworzenia nowych technik pozwalających w wiarygodny sposób przedstawić kłamstwo.

Przykładem osoby, która w znakomity sposób wykorzystwała siłę daną mu przez telewizję był Edward Bernays nazywany "Ojcem public relations". W roku 1929 został on zatrudniony aby wypromować papierosy firmy "Lucky Strike" stworzona przez niego reklama ukazywała kobiety maszerujące ulicami Nowego Yorku podczas palenia papierosów. Ponieważ kobiety palące były uznawane w tamtych czasach tematem taboo autor reklamy nazwał ją w gazetach walką o prawa kobiet. Reklama ta spowodowała tak duże spopularyzowanie palenia papierosów, że właśnie Edwardowi Bernays głównie przypisuje ich dużą sprzedaż przez kolejne lata. Był on także osobą dzięki której w dzisiejszych czasach diamenty są uznawane za symbol miłości po tym jak został zatrudniony do wypromowania diamentów firmy "De Beers" [6].

### 3.1.3. Internet

Pojawienie się internetu wpłynęło na każdy element życia codziennego czynności takie jak komunikacja, rozrywka, a także rozprzestrzenianie informacji uległy zmianom tak wielkim, że ciężko wyobrazić sobie w jaki sposób działały one wcześniej.

Dzięki pojawieniu się takich portali społecznościowych jak Facebook, Twitter oraz Instagram każdy użytkownik internetu może opowiedzieć o swoich przemyśleniach, wydarzeniach z życia każdej osobie zainteresowanej. Portale te pozwoliły nie tylko na przekazywanie informacji o sobie ale także opowiadanie o wydarzeniach ze świata przez wszystkie chętne osoby. Wraz z rozwojem internetu stopniowo zwiększa się ilość osób czerpiących informacje na temat

wydarzeń właśnie z portali społecznościowych i do dnia dzisiejszego w Ameryce wynosi 68% dorosłych osób.

Udostępnienie każdej osobie możliwości wypowiedzenia się doprowadziło do sytuacji w której duża część informacji w internecie jest całkowicie fałszywa bądź w pewien sposób zmanipulowana poprzez osobę nieobiektywnie opisującą wydarzenia. Ogrom informacji można zauważyć na podstawie ilości potwierdzonych fake newsów związanych z wydarzeniami wokół wirusa COVID-19, których do 20-06-2020 jest aż 110 niektóre z nich to [2]:

- Szczepionka na koronawirusa jest ukrywana od marca
- Aspiryna jest lekarstwem na COVID-19
- Komary przenoszą koronawirusa
- Kraje bez sieci 5G są wolne od koronawirusa
- Koronawirus nie zagraża uczestnikom zgromadzeń religijnych
- Koronawirus to środek do zmniejszenia populacji Ziemi

Osoby oszukane fake newsami grają istotną rolę w dalszym ich rozprzestrzenianiu poprzez udostępnianie informacji swoim znajomym lub rozmowy jest to cecha internetu, która daje niespotykaną wcześniej efektywność oszukiwania dużej ilości ludzi w szybkim czasie.



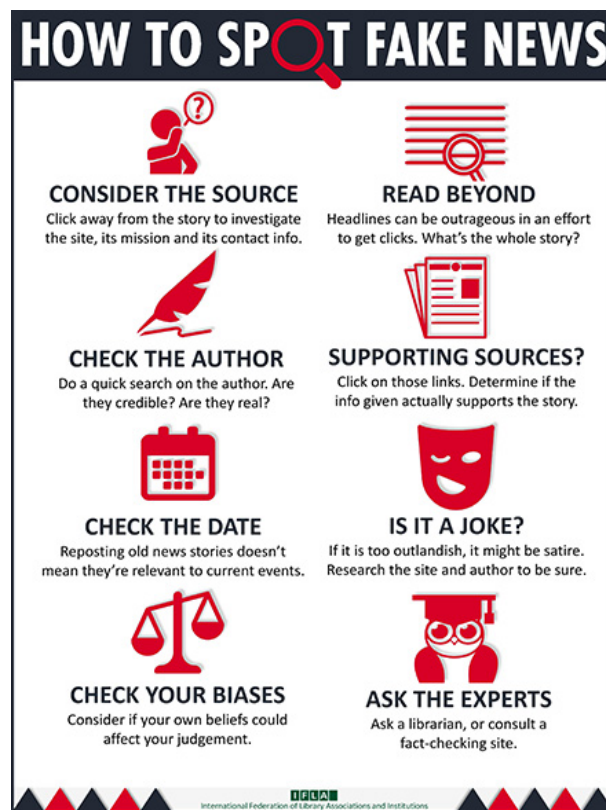
Rys. 3.3: Przykład fałszywej informacji na portalu facebook

Na powyższym obrazie ukazany jest post udostępniony na portalu Facebook z którego wynika, że jeżeli osoba jest w stanie wstrzymać oddech na 10 sekund to nie jest ona zarażona na koronawirusa. Metoda ta według opisu miała być skuteczna według ekspertów z Japonii. Jak się później okazało informacja ta była całkowicie fałszywa nie powstrzymało to jednak ponad 2.4 tysiąca ludzi przed udostępnieniem jej swoim znajomym. [5]

## 3.2. Sposoby ochrony przed nieprawdziwymi informacjami

Rozwój tak potężnego narzędzia jak internet spowodował że fałszywe informacje stały się poważnym zagrożeniem w dzisiejszym świecie. Aby zapobiec oszukaniu dużej ilości społeczeństwa znaleziono różne sposoby na ochronę przed nieprawdziwymi informacjami niektóre z nich to

1. IFLA “How to spot fake news”- jest to stworzona przez międzynarodową instytucję reprezentującą interesy bibliotekarzy i pracowników informacji o nazwie IFLA infografika przedstawiające listę rzeczy, które należy zrobić podejrzewając że jesteśmy oszukiwani.



Rys. 3.4: Infografika stworzona przez IFLA

Czynności które są na niej zawarte to

- Sprawdzenie źródła informacji
  - Dokładne przeczytanie treści
  - Sprawdzenie autora
  - Analiza odnośników
  - Sprawdzenie dat związanych
  - Upewnienie się że informacja nie jest formą żartu
  - Obiektywna ocena informacji
  - Zapytanie ekspertów
2. Strony internetowe stworzone w celu walki z dezinformacją - istnieje wiele witryn gdzie eksperci sprawdzają nadesłane przez użytkowników informacje pod względem ich zgodności z prawdą. Przykładami takich portali są: *fakenews.pl*, *snopes.com*, *FactCheck.org*, *factchecker.in*
  3. Grupy osób powołanych przez rząd do walki z fałszywymi informacjami - najlepszym przykładem takiej grupy są tzw. *Litewskie Elfy* jest to grupa ochotników, którzy w wolnym czasie przeglądają fora internetowe oraz media społecznościowe sprawdzając znajdujące

się na nich informacje. W przypadku znalezienia nieprawdziwej informacji osoby te informują administratora strony o problemie co najczęściej prowadzi do jego rozwiązania. Nazwa grupy wzięła się z faktu że walczą oni z grupami powszechnie zwanymi *trollami*. Grupa ta zyskała na popularności w takim stopniu że rozpoczęto organizację kolejnych oddziałów w innych krajach między innymi na Łotwie. [3]

Pomimo istnienia wielu sposobów ochrony przed fałszywymi informacjami ich ilość powoduje, że bardzo łatwo zostać oszukanym. Z tego powodu bardzo pomocne byłoby stworzenie oprogramowania pozwalającego na automatyczne rozpoznanie czy informacja jest prawdziwa. Implementacja takiego systemu w mediach społecznościowych jak *facebook*, *twitter* pozwoliłoby na usunięcie kłamstw zanim trafiłyby one przed oczy użytkowników. Popularne w ostatnich czasach algorytmy uczenia maszynowego *ML* oraz analizy języka naturalnego *NLP* osiągają zaskakująco dobrą efektywność w klasyfikacji różnego rodzaju tekstów więc ich wykorzystanie w rozwiązaniu takiego problemu mogłoby pozwolić na rozwiązanie go albo zmniejszenie.

### 3.3. Deep fake

# **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

- [1] Fake news. it's complicated. <https://firstdraftnews.org/latest/fake-news-complicated/>. Dostęp dnia: 15-06-2020.
- [2] Krótki przewodnik po fake newsach o koronawirusie. <https://www.cyberdefence24.pl/krotki-przewodnik-po-aktualnych-fake-newsach-o-koronawirusie>. Dostęp dnia: 15-06-2020.
- [3] Obrona przestrzeni informacyjnej na przykładzie litwy, Łotwy i estonii. <https://warsawinstitute.org/pl/obrona-przestrzeni-informacyjnej-na-przykladzie-litwy-lotwy-estonii/>. Dostęp dnia: 15-06-2020.
- [4] Słownik języka polskiego. <https://sjp.pwn.pl/>. Dostęp dnia: 15-06-2020.
- [5] Wstrzymaniem oddechu nie sprawdzisz, czy masz koronawirusa. [https://demagog.org.pl/analizy\\_i\\_raporty/wstrzymaniem-oddechu-nie-sprawdzisz-czy-masz-koronawirusa/](https://demagog.org.pl/analizy_i_raporty/wstrzymaniem-oddechu-nie-sprawdzisz-czy-masz-koronawirusa/). Dostęp dnia: 15-06-2020.
- [6] M. Dice. *The True Story of Fake News: How Mainstream Media Manipulates Millions*. The Resistance Manifesto, 2017.
- [7] F. L. Mott. *American Journalism*. Macmillan, 1941.

## **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.