FAKE NEWS DETECTION

AI PRODUCT SERVICE

BUSINESS/FINANCIAL MODELLING

By

Satyanarayan Sahoo, Madhavi Parashar, Akash Valluri

26th June 2023

Abstract:

This paper presents a business idea for a machine learning-based product that can detect fake articles and spam content on the internet. The product will utilize state-of-the-art natural language processing techniques to analyze text data and identify patterns and features that are indicative of spam or fake content. The product will be aimed at businesses and individuals who need to protect their online reputation and prevent the spread of misinformation.

The proposed product will be built using advanced machine learning algorithms, including deep learning neural networks, to achieve high accuracy in detecting fake articles and spam content. The paper outlines the potential market size for this product and the potential revenue streams, including subscription-based models and partnerships with social media platforms and online news outlets. Overall, this business idea has significant potential to meet the growing demand for effective spam and fake article detection in the digital age.

1. Problem statement:

The increasing volume of online content has led to a rise in the spread of fake news and spam, which can harm businesses and individuals by promoting fraudulent or low-quality products or services. Spam can harm the reputation of a company or product, reduce productivity, and waste resources. Therefore, detecting and removing spam is critical for maintaining a healthy and efficient system.

Using a business idea for a fake article or spam detection using a machine learning model can solve a variety of problems, including: Using machine learning can **prevent fraud** by detecting and filtering out fake articles and spam that promote fraudulent products or services.

1. Filtering out fake articles and spam *can **improve the customer experience*** by reducing irrelevant and misleading content.

1. Businesses can **enhance their brand reputation** by taking steps to filter out fake articles and spam, which can damage their reputation and erode customer trust.

2. Machine learning can help businesses **identify trends and opportunities** by analysing data from articles and other sources, which can inform strategic decision-making.

3. Automating the filtering process using machine learning can **save time and resources** that would otherwise be spent manually filtering out fake articles and spam.

a. Feasibility: The Fake News Detection Software solution can be developed within the next 2 years, considering the availability of existing technologies and expertise in Machine Learning and AI.

b. Viability: The Fake News Detection Software solution addresses the long-term challenges faced by the online/social media industry in receiving right and authenticate information, making it relevant and capable of surviving for a long duration of 3-4 decades and growing continuously as the advances in the IT sector increases with time.

c. Monetization: The Fake News Detection Software solution can be directly monetized by offering a subscription-based service to social media platforms, online fact-checkers, regular users having an own user-interface service. This ensures direct revenue generation.

2.Final Product Prototype

2.1 Abstract:

Our implementation contains 3 search fields which are

1) Search by article content.

2) Search using key terms.

3) Search for website in database.

In the first search field we have used Natural Language Processing for the first search field to come up with a proper solution for the problem, and hence we have attempted to create a model which can classify fake news according to the terms used in the newspaper articles. Our application uses NLP techniques like Count Vectorization and TF-IDF Vectorization before passing it through a Passive Aggressive Classifier to output the authenticity as a percentage probability of an article.

The second search field of the site asks for specific keywords to be searched on the net upon which it provides a suitable output for the percentage probability of that term actually being present in an article or a similar article with those keyword references in it. The third search field of the site accepts a specific website domain name upon which the implementation looks for the site in our true sites database or the blacklisted sites database. The true sites database holds the domain names which regularly provide proper and authentic news and vice versa. If the site isn't found in either of the databases then the implementation doesn't classify the domain it simply states that the news aggregator does not exist.

Working:

Data pre-processing:

- Collect a large dataset of articles, including both genuine and fake/spam articles.
- Clean the data by removing any HTML tags, URLs, or other non-textual elements.
- Tokenize the text into individual words or subworlds.
- Normalize the text by converting everything to lowercase and removing punctuation.
- Remove stop words, which are common words that don't carry much meaning (e.g., "the", "and", "a").

Text pre-processing:

- Convert the text data into numerical representations, such asword embeddings or bag-of-words representations.
- Split the data into training and testing sets.

Model selection:

- Choose a machine learning model that is suitable for text classification, such as Naive Bayes, logistic regression, or a neural network. We used Multinomial NB classifier and Passive Aggressive classifier.
- Tune the hyperparameters of the model using techniques such as grid search or random search.

Model building:

- Train the machine learning model on the training set.
- Evaluate the model's performance on the testing set using metrics such as accuracy, precision, recall, and F1 score.
- If the model's performance is not satisfactory, try tweaking the model architecture or hyperparameters and retrain the model.

Deployment:

Deploy the machine learning model to a production environment, where it can be used to classify new articles as genuine or fake/spam.
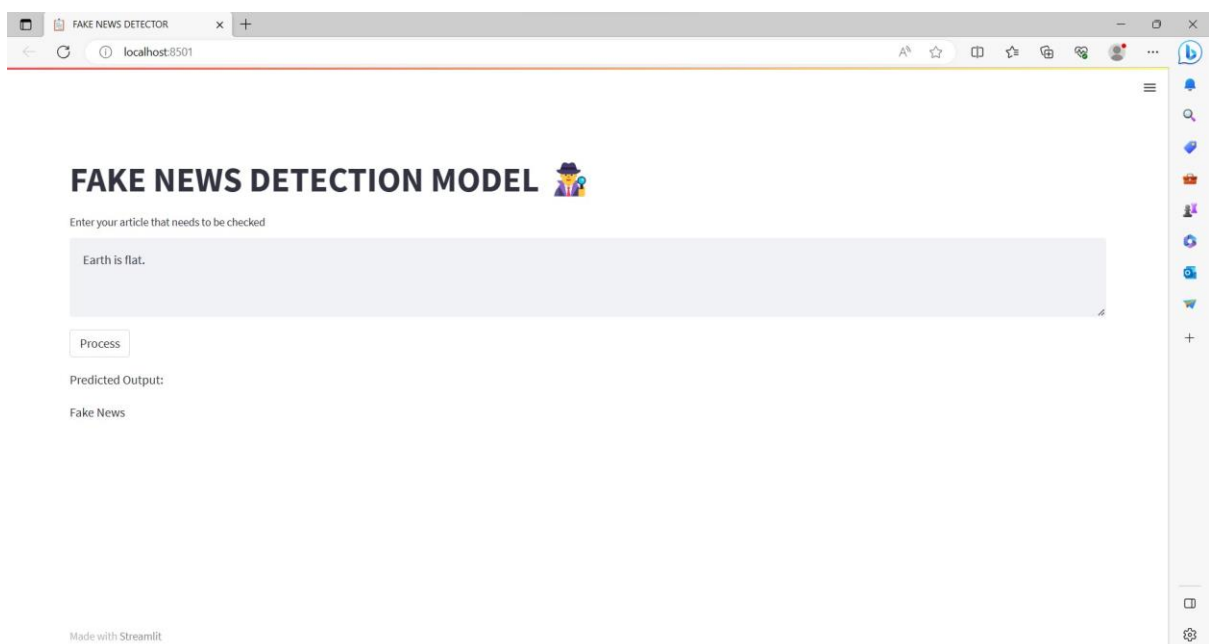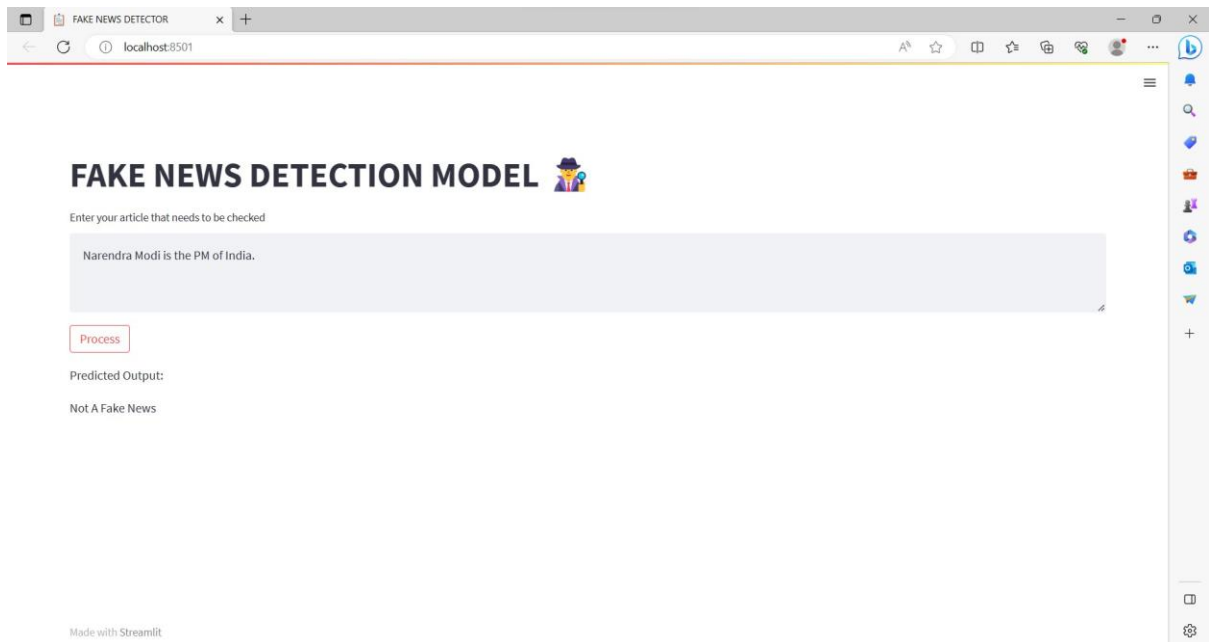
You can test it yourself in the following link:
https://fakenewsdetector-p1sd2r9abab.streamlit.app/

Full code implementation in this link:
https://github.com/SatyaTheG/Fake_news_detector

Algorithm:

The problem can be broken down into 3 statements

1. Use NLP to check the authenticity of a news article.

2. If the user has a query about the authenticity of a search
    query, then we he/she can directly search on our platform and
    using our custom algorithm we output a confidence score.

3. Check the authenticity of a news source. These sections have
    been produced as search fields to take inputs in 3 different forms
    in our implementation of the problem statement.

Some Basic Visualizations on Real World or Augmented Data:

Top 5 websites posting Real News

With stopwords

Without stopwords
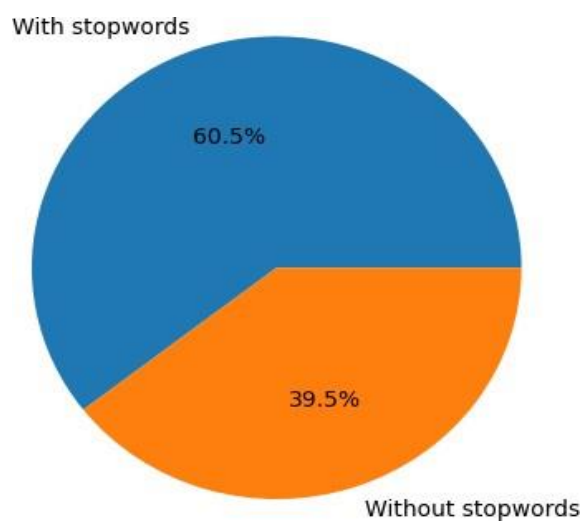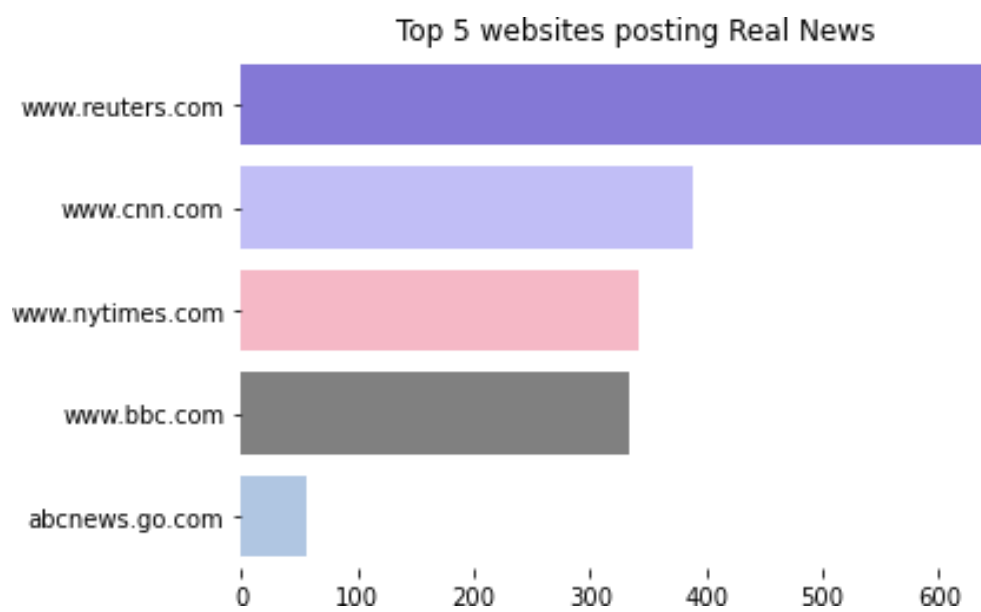
BUSINESS MODEL

Monetization Strategies: Different revenue streams for business model, such as:

- Subscription Model: Offer different pricing tiers for individuals, organizations, or media outlets to access fake news detection service.
- Licensing and Partnerships: Collaborate with social media platforms, news agencies, or fact-checking organizations to provide them with your detection technology through licensing agreements.
- API Access: Charge a fee for providing programmatic access to fake news detection API, allowing other businesses to integrate solution into their platforms.
- Consultancy and Training: Offer consulting services or training workshops to educate organizations and individuals on identifying and combating fake news.
- Marketing and Outreach: Develop a marketing strategy to raise awareness about our solution. Leverage social media, content marketing, and collaborations with media outlets or fact-checking organizations to establish credibility and attract customers.
- Legal and Ethical Considerations: Familiarize yourself with legal and ethical considerations related to privacy, data protection, and bias mitigation. Ensure that business model adheres to applicable regulations and guidelines.
- Remembering that developing an effective fake news detection model is a challenging task, and it requires continuous improvement and adaptation. Collaborating with experts in the field, fact-checking organizations, and conducting independent evaluations can help enhance the credibility of solution.

(A) MARKET ANALYSIS:

The product/service of Fake News Detection using Machine Learning can be launched into the technology market, specifically within the domain of online platforms, news aggregators, social media networks, and digital media companies. These markets are particularly relevant due to the growing concerns surrounding the spread of misinformation and the need for effective tools to combat fake news.

By targeting these markets, the product/service aims to provide a solution that assists in identifying and flagging false or misleading information, helping users make informed decisions and promoting the credibility of online content. The technology market is continuously evolving, and there is a demand for innovative solutions that can enhance trust and accuracy in information dissemination.

It's important to note that while the product/service can be launched into the technology market, it may also have implications in the media industry, journalism sector, and regulatory bodies concerned with maintaining the integrity of information. Therefore, collaboration with these stakeholders could be beneficial to ensure wider adoption and impact.

(B) OPERATING PLAN:

The operating plan for the business model of a website used for Fake News Detection Using Machine Learning involves several key components. Firstly, the website would require a user-friendly interface that allows users to submit news articles or URLs for analysis. The website would then utilize machine learning algorithms to analyse the content and detect any signs of fake news based on various factors such as credibility of the source, linguistic patterns, and fact-checking databases.

To ensure accuracy and reliability, the website would need to continuously update its machine learning models by incorporating new data and information. This would involve employing a team of data scientists and researchers who can constantly monitor and improve the algorithms.

The website can generate revenue through several channels. One approach is to offer a freemium model, where basic usage is free, but advanced features such as detailed analysis reports or real-time monitoring require a subscription fee. Additionally, partnerships can be established with news organizations, social media platforms, or online advertising networks to provide them with access to the fake news detection service for a fee.

To attract users and gain credibility, the website would need to invest in marketing and outreach efforts, collaborating with fact-checking organizations, academic institutions, and media watchdogs to promote the service. Building a strong brand reputation and establishing trust among users would be crucial for the long-term success of the business model.

Overall, the operating plan would focus on developing a robust and accurate fake news detection system, monetizing through subscriptions and partnerships, and actively engaging with stakeholders to promote the importance of combating fake news in today's digital landscape.

(C) MARKETING PLAN:

Our marketing plan for the business model of our website, which focuses on Fake News Detection Using Machine Learning, aims to raise awareness about the prevalence of fake news and provide a reliable solution for users. Firstly, we will employ a multi-channel approach to reach our target audience, including online advertisements, social media campaigns, and partnerships with reputable news outlets and organizations. Through these

channels, we will emphasize the importance of verifying information and showcase our website as a trustworthy tool for identifying fake news. Additionally, we will leverage content marketing by creating informative blog posts, videos, and infographics to educate users about the dangers of fake news and the effectiveness of our machine learning algorithms. By positioning our website as a go-to resource for reliable news, we will attract a substantial user base and establish ourselves as a reputable brand in the fight against misinformation.

D.FINANCIAL EQUATION:

To develop a mathematical financial equation for a website that provides Fake News Detection using Machine Learning, we need to consider various factors such as revenue sources, expenses, and subscriber base. Here's an example equation using a specific number of subscribers:

Let's assume the following variables:

S = Number of subscribers

P = Monthly subscription price

R = Monthly revenue from subscriptions

C = Monthly operating costs

I = Monthly investment in technology and infrastructure

E = Monthly expenses (excluding investment)

The equation can be expressed as follows:

Monthly Net Profit = (R - C - I) - E

To calculate the monthly revenue from subscriptions (R), we can use the equation:

R = S * P

This equation multiplies the number of subscribers (S) by the monthly subscription price (P) to determine the total monthly revenue from subscriptions.

Please note that this equation is a basic model and doesn't take into account other factors like advertising revenue, partnerships, or additional services. It assumes that the primary revenue source is the subscription fees from the subscribers. Additionally, the equation

doesn't consider factors such as churn rate (subscriber attrition) or acquisition costs, which could affect the overall financial performance of the business model.

Remember that real-world financial models are more complex and may require additional variables and considerations based on the specific context and dynamics of the business.

Conclusion:

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hard-copies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. WhatsApp's forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed. Thus, in order to curb the phenomenon, we have developed our Fake news Detection system that takes input from the user and classify it to be true or fake. To implement this, various NLP and Machine Learning Techniques have to be used.