PARSHVANATH CHARITABLE TRUST'S
A. P. SHAH INSTITUTE OF TECHNOLOGY
Department of Information Technology
(NBA Accredited)

Hate Speech Detection & Fake News Detection

Sakshi Balekar **20104103**
Sarthak More **20104116**
Prathamesh Lambate **20104064**

Project Guide
Ms. Sonal Jain

Contents

1. Introduction

- The widespread use of social media platforms, which has made it easier for hate speech and fake news to spread quickly and reach a large audience.
- The negative impact that hate speech and fake news can have on individuals and society as a whole, including inciting violence, spreading misinformation, and undermining trust in institutions.

Problem Identified:

- They can create echo chambers and reinforce existing biases, making it harder for people to understand different perspectives and engage in productive dialogue.
- They can be difficult to identify and address, especially on social media platforms, where content can spread quickly and reach a large audience.

Solution Proposed:

- Developing and implementing effective content moderation policies and tools on social media platforms, including automated systems and human review processes.
- Promoting media literacy and critical thinking skills to help individuals identify and evaluate sources of information.
- So, here we have come up with a concept of HATE SPEECH DETECTION AND FAKE NEWS DETECTION.

2. Objectives

- To analyze online content and identify instances of hate speech and fake news.
- To measure the effectiveness of detection algorithms.
- To promote greater media literacy, by encouraging critical thinking and informed decision-making among online users.
- To develop detection methods that are effective, ethical, and sustainable in the long-term.
- To automatically analyse text and identify potentially problematic content.

3. Scope

- Can be integrated into social media platforms to automatically identify and remove harmful or false content.
- Can use these technologies to fact-check and verify news articles, ensuring that they are accurate and free from misinformation.
- Can use these technologies to teach media literacy and critical thinking skills to students, helping them to evaluate sources of information and identify fake news.
- Can use these technologies to monitor and analyze customer feedback and social media mentions, helping them to identify potential issues and respond to negative publicity.

4. Literature Survey

STUDY TITLE	RESEARCH QUESTION	DATASET	ALGORITHM USED	KEY FINDINGS
"Automated Hate Speech Detection and the Problem of Offensive Language" by Davidson et al. (2017)	Can machine learning algorithms accurately detect hate speech?	Twitter dataset	SVM classifier with various feature sets	Achieved 91% accuracy in detecting hate speech, but struggled with identifying offensive language that was not explicitly hateful.
"Fake News Detection on Social Media: A Data Mining Perspective" by Shu et al. (2017)	How can fake news be automatically detected on social media?	Twitter dataset	SVM and Random Forest classifiers with linguistic and user-based features	Found that user-based features were more effective than linguistic features in detecting fake news.
"Detecting Misinformation and Fact-checking on Social Media: A Data Mining Perspective" by Jin et al. (2020)	Can machine learning algorithms distinguish between misinformation and factual news on social media?	Weibo dataset	BiLSTM and CNN models with textual and social context features	Achieved 90% accuracy in distinguishing between misinformation and factual news, with social context features being particularly effective.
"Multimodal Hate Speech Detection using Deep Learning" by Mandal and Bandyopadhyay (2020)	How can multimodal features (e.g., text and images) improve hate speech detection?	Gab and Twitter datasets	BiLSTM and CNN models with both textual and visual features	Found that incorporating visual features improved hate speech detection, especially for images with explicit content.
"Detecting Propaganda Techniques in News Articles: A Deep Learning Approach" by Barrón-Cedeño et al. (2019)	How can deep learning models be used to automatically detect propaganda techniques in news articles?	Propaganda dataset	BiLSTM and attention models with textual features	Achieved 73% accuracy in detecting propaganda techniques, with attention models outperforming BiLSTM models.

5. Proposed System

- Can upload the document containing hate speech or fake news according to the need.
- Can upload the text data containing irrevelant/hate content.
- Can convert the image to text and then detect the error.
- Can detect the news and tell us if it is fake or not.
- Can detect voice and convert it into the text.

6. Algorithms Used

Decision Tree:

- The first step in applying decision tree algorithm to hate speech and fake news detection is to preprocess the data.
- This involves cleaning the text data, removing any irrelevant information, and converting the text data into a numerical format that can be used by the algorithm.
- After pre-processing the data, the next step is to select relevant features that can help the algorithm distinguish between hate speech or fake news and non-hate speech or non-fake news.
- This may involve using techniques such as bag-of-words, word embeddings, or other feature extraction techniques.
- Once the relevant features have been selected, the next step is to train the decision tree on a labeled dataset of examples.
- The decision tree will learn to classify the input text based on the features that have been selected.

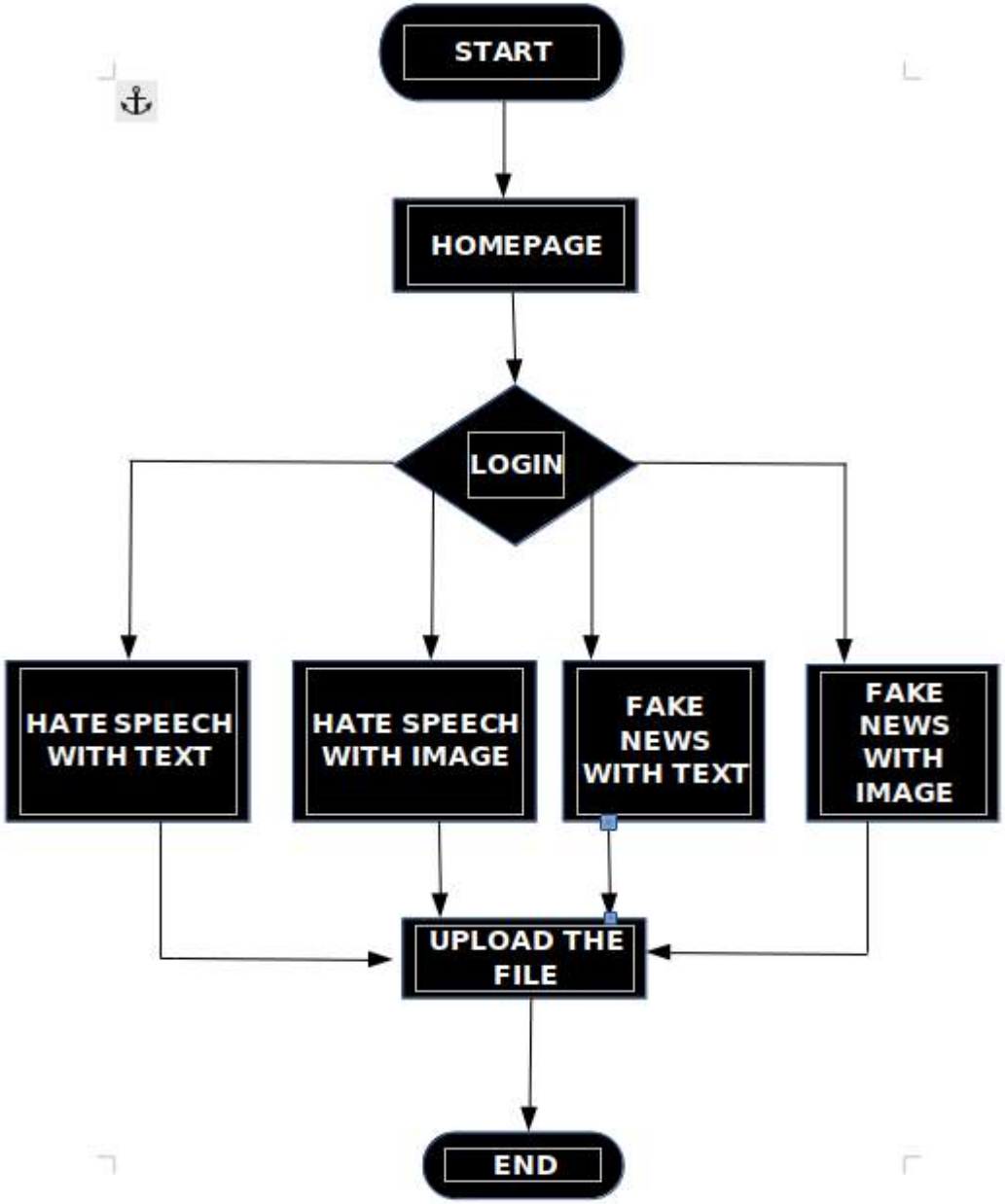
- After training the decision tree, it is important to evaluate its performance on a separate test dataset.
- This will give an idea of how well the decision tree can generalize to new data.
- If the performance of the decision tree is not satisfactory, it may be necessary to fine-tune the algorithm by adjusting its parameters or adding more features to the model.

7. Outcome of Project

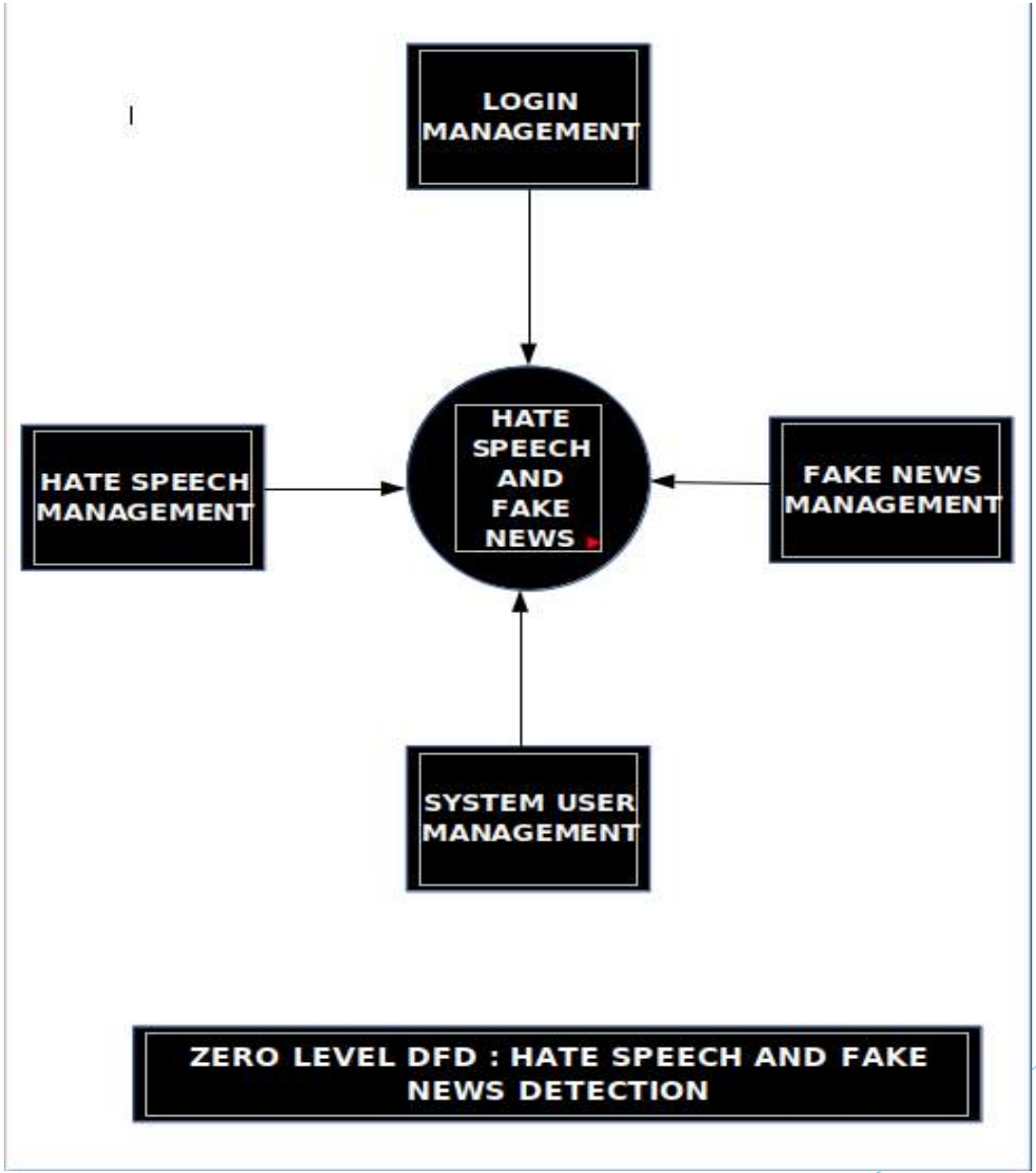
- The system should be able to identify harmful language and discriminatory content in social media posts, news articles, and other sources.
- It should be able to distinguish between hate speech and protected speech, such as political speech or satire.
- The system should be able to identify misinformation and false information in the news media, social media, and other sources.
- It should also be able to provide context and correct information to counter the false information.

- User can use the system to classify incoming data as hate speech or fake news.
- The system can be used to monitor social media platforms, news websites, and other online sources to detect hate speech and fake news.
- User can use the system to moderate content and remove hate speech and fake news before it is published.
- User can use the system to analyze and report on the trends and patterns of hate speech and fake news in the data.

8. Block Diagram



9. Data Flow Diagram



10. Technology Stack

Frontend:

- Streamlit

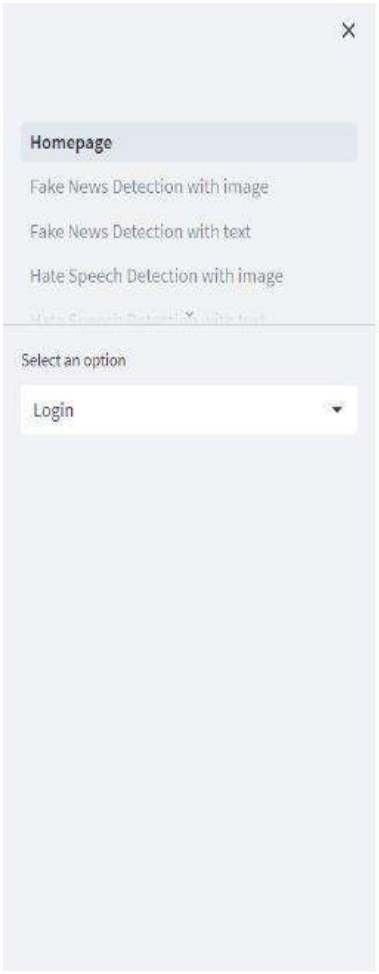
Backend:

- Python Using ML Algorithms

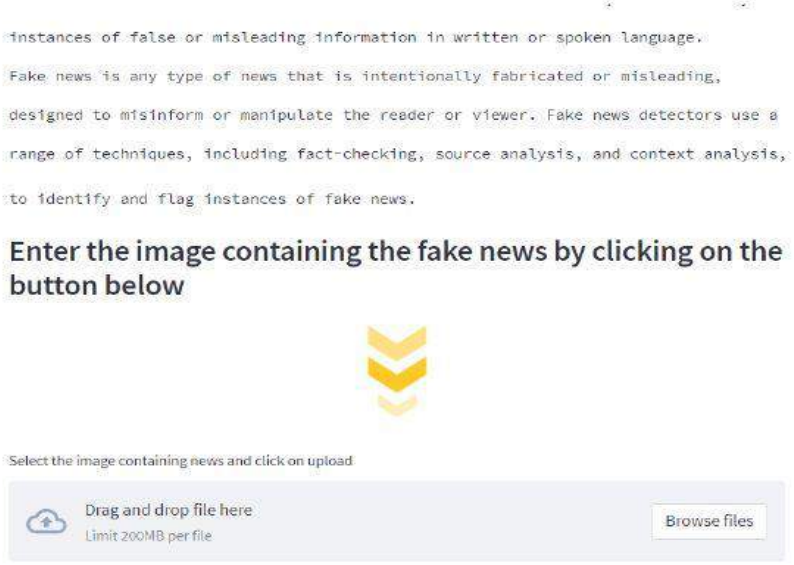
11. Suggestions in Review-1

- Write the literature survey in the tabular format.
- Use only a single algorithm instead of many algorithms.
- Select the algorithm with the best accuracy.
- The user should also be able to upload an image, then the algorithm is supposed to first convert it to text and then detect the hate speech and fake news.
- Change the block diagram.

12. Result and Discussion




Homepage
Fake News Detection with image
Fake News Detection with text
Hate Speech Detection with image
THE TRUTH PROJECT
THE TRUTH PROJECT
Hate Speech and Fake News Detector.


instances of false or misleading information in written or spoken language.
Fake news is any type of news that is intentionally fabricated or misleading,
designed to misinform or manipulate the reader or viewer. Fake news detectors use a
range of techniques, including fact-checking, source analysis, and context analysis,
to identify and flag instances of fake news.
Enter the image containing the fake news by clicking on the button below
Select the image containing news and click on upload
Drag and drop file here
Limit 200MB per file
Browse files


Homepage
Fake News Detection with image
Fake News Detection with text
Hate Speech Detection with image
Hate Speech Detection with text
Hate Speech Detection with Voice
THE TRUTH PROJECT
THE TRUTH PROJECT
Hate Speech and Fake News Detector.


The Truth Project
HATE SPEECH
Hate Speech Detector
Enter the image containing the hate speech by clicking on the button below
Drag and drop file here
Browse files


Homepage
Fake News Detection with Image
Fake News Detection with text
Hate Speech Detection with Image
Hate Speech Detection with text
Hate Speech Detection with Voice
THE TRUTH PROJECT
THE TRUTH PROJECT
Hate Speech and Fake News Detector.


The Truth Project
The Truth Project is a software tool that uses artificial intelligence and natural language processing techniques to identify instances of hate speech in written or spoken language.
HATE SPEECH
Hate Speech Detector
Hate speech is any form of speech that targets a particular group of people based on their race, ethnicity, religion, gender, sexual orientation, or other characteristics. Hate speech detectors use a range of techniques, including keyword analysis, sentiment analysis, and context analysis, to identify and flag instances of hate speech.
Enter the text containing the hate speech in the input section below
Please login to continue


Homepage
Fake News Detection with Image
Fake News Detection with text
Hate Speech Detection with Image
Hate Speech Detection with text
Hate Speech Detection with Voice
THE TRUTH PROJECT
THE TRUTH PROJECT
Hate Speech and Fake News Detector.


The Truth Project
The Truth Project is a software tool that uses artificial intelligence and natural language processing techniques to identify instances of hate speech in written or spoken language.
HATE SPEECH
Hate Speech Detector
Hate speech is any form of speech that targets a particular group of people based on their race, ethnicity, religion, gender, sexual orientation, or other characteristics. Hate speech detectors use a range of techniques, including keyword analysis, sentiment analysis, and context analysis, to identify and flag instances of hate speech.
Click on the button below to give voice as an input
Press and Speak

13. Conclusion and Future Scope

In conclusion, hate speech and fake news are critical issues that can cause harm to individuals and society. The detection of hate speech and fake news is an important area of research, and the development of automated systems that can identify such content can help to prevent harm and protect individuals and communities.

The data flow diagram and use case diagram provide an overview of the components and functionality of a hate speech detection and fake news detection system from both user and developer perspectives. The entity relationship model and block diagram further illustrate the data structures and system architecture.

Looking to the future, there are several areas of research and development that can further improve the accuracy and efficiency of hate speech detection and fake news detection systems. For example, the use of more sophisticated natural language processing algorithms and machine learning models can improve the ability to detect and classify different types of hate speech and fake news content.

14. References

Here are some useful references and links for hate speech detection and fake news detection:

- "A Survey on Hate Speech Detection using Natural Language Processing" by Pooja Sharma and Sumit Pandey, International Journal of Advanced Research in Computer Science, Volume 8, No. 4, 2017. https://www.ijarcs.info/index.php/Ijarcs/article/view/3265
- "Fake News Detection on Social Media: A Data Mining Perspective" by S. Arora, A. Gupta, R. Gupta, and P. Kumar, ACM SIGKDD Explorations Newsletter, Volume 19, Issue 1, 2017. https://dl.acm.org/doi/10.1145/3178042.3178056
- "Hate Speech Detection: A Solved Problem?" by Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber, Proceedings of the First Workshop on Abusive Language Online, 2017. https://www.aclweb.org/anthology/W17-3003/
- "Fake News Detection: A Deep Learning Approach" by N. Shah and A. Kumar, 2018 IEEE 8th International Advance Computing Conference (IACC), 2018. https://ieeexplore.ieee.org/abstract/document/8691816
- "Hate Speech Detection and Analysis Using Machine Learning and Deep Learning Techniques: A Review" by S. Malik, A. Kumar, and R. Aggarwal, 2019 IEEE 9th International Advance Computing Conference (IACC), 2019. https://ieeexplore.ieee.org/abstract/document/8971329

Thank You...!!