International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

Hate Speech Detection in social Media Using Deep Learning

[1]Nerella Vivek, [2]Dr. Ramavtar

[1]*Research Scholar, Dep of Computer Science and Engineering, Glocal University.*
[2]*Assistant Professor, Dep of Computer Science and Engineering, Glocal University.*

Date of Submission: 28-02-2024 Date of Acceptance: 08-03-2024

ABSTRACT
YouTube, a popular platform for global dialogue, has become a hub for hate speech, which can undermine community trust, propagate harmful ideologies, and even incite violence. Traditional methods of detecting and moderating hate speech on YouTube involve a combination of automated systems and human moderators. However, these methods face challenges such as the volume and scale of the platform, which makes manual review impractical and automated systems prone to errors. Hate speech often involves subtle and context-dependent language, making it difficult for algorithms to discern. Additionally, the language used in hate speech evolves rapidly, including coded or euphemistic language that can evade detection by simpler keyword-based systems. Therefore, a more comprehensive approach is needed to effectively manage and moderate hate speech on YouTube.

KEYWORDS: Youtube, Deep Learning, Hate Speech and Machine Learning.

I. INTRODUCTION

Digital images, videos, and other visual inputs may have information extracted from them by computers and systems that are equipped with computer vision, a branch of artificial intelligence. Conclusions or suggestions may be drawn from this data. While artificial intelligence (AI) enables computers to carry out cognitive activities, computer vision allows computers to analyses and comprehend visual data. Computer vision certain advantages to having good eyesight due to the complexity of the human visual system. One of the most impressive features of the human visual system is its ability to continually absorb new information. It lets us distinguish between objects, measure distances, detect motion, and spot abnormalities in images. By combining data, algorithms, and cameras, computer vision makes it possible for computers to perform a wide range of tasks once performed by biological parts of the brain, including the visual cortex, optic nerves, and retinas. One advantage of automated monitoring and inspection systems is that they can evaluate more objects or processes per minute than human inspectors can. Their talent for spotting hidden issues is unparalleled. The energy, utility, manufacturing, and automotive industries are just a few that rely heavily on computer vision technology. The computer vision sector is also growing at a rapid pace. Picture classification, object detection, object tracking, and action recognition are all subfields of computer vision (fig. 1).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

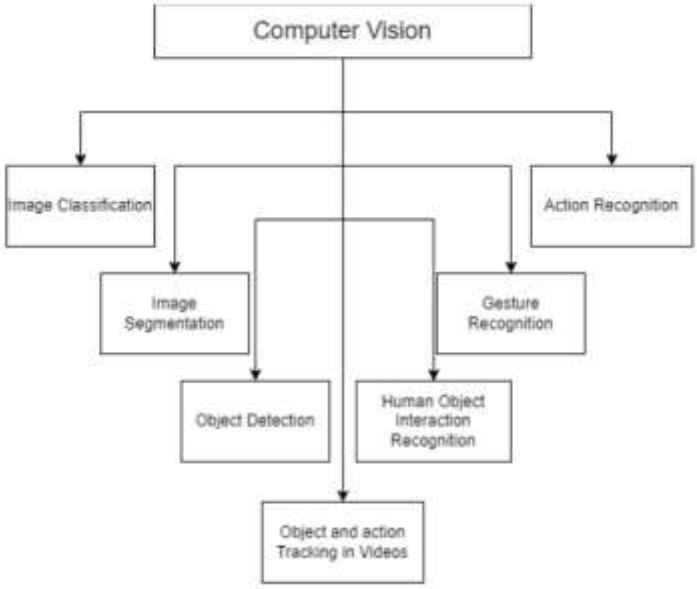

Fig. 1: Various dimensions of computer vision

II. THE ROLE OF DEEP LEARNING

Deep learning, a powerful tool that uses multi-layer neural networks, has emerged as a powerful solution for identifying hate speech in YouTube videos. These models can learn from large datasets and spot patterns and contextualized signals, making them capable of understanding the nuances of language, including the exact context of words and sentences. They are also capable of scalability, as the increasing data volume on sites like YouTube is no match for these algorithms, which can be trained on massive datasets. This research aims to improve hate speech identification in YouTube videos by developing deep learning models that can comprehend the complex nature of hate speech. The goal is to compare their accuracy, precision, and memory to more conventional approaches. The project proposes practical strategies for integrating these models into existing moderation systems on YouTube, ensuring they complement human oversight and improve overall content safety. Improving YouTube's capacity to detect hate speech using powerful deep learning technologies is critical for promoting a more inclusive and safer online environment. This project aims to develop and use models that are more accurate and responsive to context. It is critical to enhance content moderation methods and tactics since internet platforms significantly impact public discussion. A more fair and polite online community might be possible with the help of deep learning, which provides a potential solution to the problems of hate speech identification. Hate speech can cause psychological harm, perpetuate stereotypes, and incite violence. Effective detection is crucial for protecting users and maintaining a respectful online environment. Current detection methods include keyword-based filtering, rule-based systems, and human moderation, which are effective but resource-intensive. Advancements in deep learning offer a state-of-the-art solution for identifying hate speech by eliminating the shortcomings of older approaches. Transformers belonging to the deep learning model family excel in grasping contextual and language subtleties, with their amazing accuracy in identifying hate speech being a testament to their immense knowledge. The proposed approach involves data collection and annotation, multimodal data analysis, textual analysis, audio analysis, and multimodal integration. Performance metrics such as F1-score, recall, accuracy, and precision are used to evaluate how well models detect hate speech. The system integration involves developing strategies for integrating these advanced models into YouTube's existing content moderation infrastructure, while human oversight ensures automated systems are complemented by human moderators who can handle complex cases and provide final judgments. Challenges and considerations include privacy and ethics, bias and fairness, and adaptation and evolution. By continuously updating models to handle emerging hate speech trends and language changes, this research aims to contribute to a more inclusive and safer online environment.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

III. TYPES OF VIDEO SUMMARIZATION TECHNIQUES

Video summarization techniques include static and dynamic methods. Static video summarization uses a set of key-frames of a video, free from time and sequence issues. Dynamic video summarization uses small portions of audio and video to sharpen the summary. This technique is useful in internet browsing and navigation, as it helps users locate the exact video and saves important messages. Video abstraction, also known as static video summarization, is formed from longer videos and creates a shot summary of the content. It can be produced using manual or automatic methods. Manual methods require more manpower, while automated techniques reduce this requirement. In still image abstract, the collection of noticeable images taken or created from the video source is called a static storyboard. The technique involves distinguishing shots, grouping them based on likeness, and positioning them for inclusion in the static summary. The static summary is created by consolidating thumbnail pictures of the chosen shots.

Dynamic video summarization, also known as moving image abstract, is a collection of images with corresponding audio extraction from the video clip with shorter length, also known as video skimming. Static video summarization uses pictorial data for fast access and no need for audio and textual data. It can be displayed in a spatial order for quick access. There are two options for summarizing a video: careful selection of original movie frames for static summarization and dynamic summarization, which incorporates both audio and motion. These qualities make the summary easier to understand and provide more useful information. Research shows that watching videos is more interesting and entertaining than using key-frames in PowerPoint presentations. Key-frames remain unaffected by timing or synchronization issues, and summaries may improve key-frame navigation in video skims. The static video summarization is illustrated in Figure 2 and that of dynamic video summarization is shown in Figure 3.

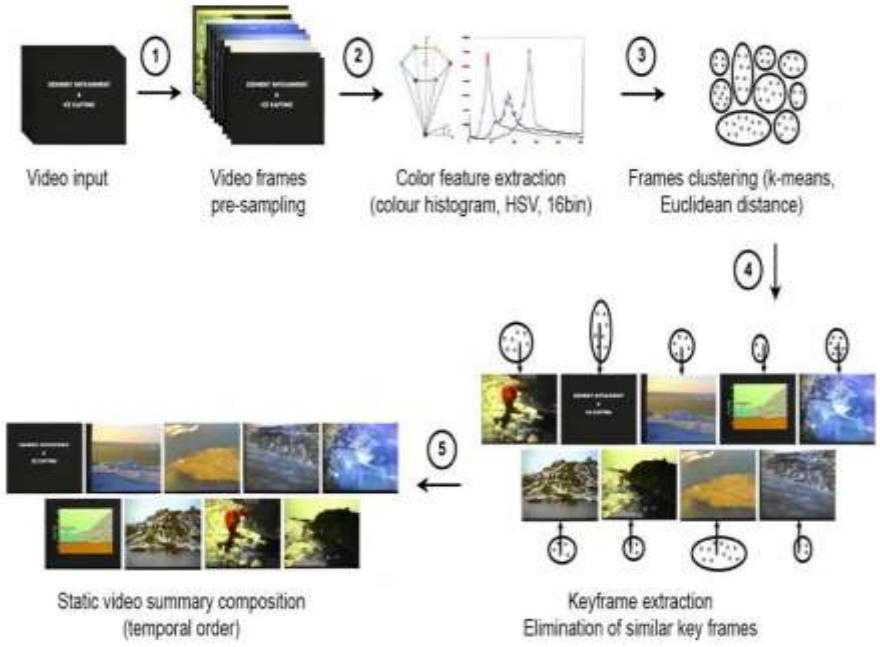

Figure 2 Static Video Summarization Process

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

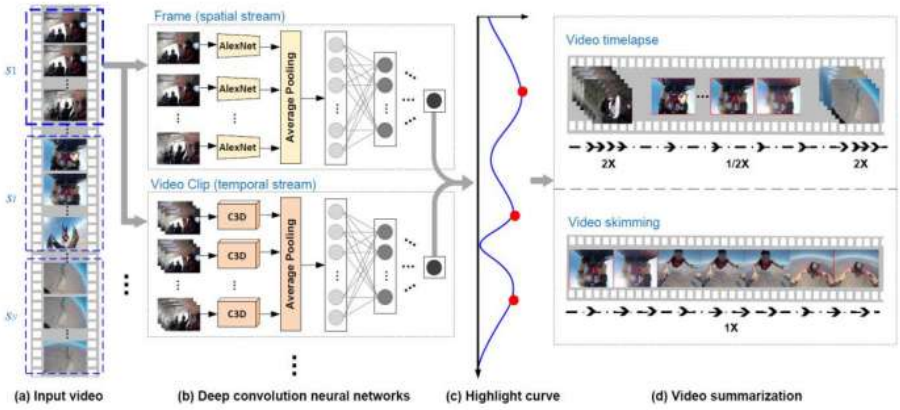

Figure 3 Dynamic Video Summarization Process

Dynamic video summarization offers several benefits, including recovering original audio information in education and training videos. However, it requires higher computational effort during the abstracting process due to playback time. The effectiveness of video summarization depends on different models and their selection. The progression of computerized mixed media has expanded the amount of advanced content available, such as films, sports, news, TV programs, home recordings, reconnaissance recordings, instructive recordings, and clinical recordings. Human clients may not have sufficient opportunity to see the entire video or have an enthusiasm to watch the whole video. In reconnaissance cameras, editors need time-sensitive video edited compositions or item-based video abstracts but do not have the enthusiasm to watch the full video. Today, even in all TV programs and home recordings need relating trailers for ad purposes. Automatic video summarization is a pressing issue to be solved, saving storage resources and time of browsing videos. Video recording is becoming less expensive and increasingly convenient with the upgrade of storage systems and fast internet speed. Automatic video summarization is a pressing issue to be solved, as it saves storage resources and time of browsing videos. The alarming rise of hate speech on social media platforms has led to significant funding for academic research on hate speech detection systems from various groups and governments. Assessing the efficacy of certain tactics may be difficult due to the distinct benefits and drawbacks associated with each potential solution. The pursuit of enhancing classification outcomes via the integration of many classifiers' capabilities is a worthwhile endeavor. Speech is considered one of the most basic and essential ways of communication in the natural world. Language symbols, which facilitate human communication, arose throughout human evolution. The ultimate goal of voice recognition technology is to convert spoken words into text that can be understood by computers. However, the growing requirements of computer-human interaction cannot be met by only relying on speech recognition to extract individual words. In this article, the authors discuss the mechanisms for understanding spoken language and propose a model that integrates a multilingual deep neural network (DNN). They use MATLAB simulations to compare and contrast two English voice recognition algorithms, GMM-HMM and CNN-CTC. The Deep Autoencoder (DAE) is a modified version of the autoencoder that incorporates a multi-layer encoder, replicating the input like a proficient expert managing data. The autoencoder exhibits a minimum reconstruction error when presented with samples from the training class, but a substantial reconstruction error is shown when the other class is offered. The Stacked De-noising Autoencoder (SDAE) is a proficient combination of denoising autoencoders that exhibit exceptional proficiency in processing YouTube footage as input. Applying feature reconstruction via the SDAE may significantly reduce the quantity of YouTube video in the input. The denoising autoencoders of the SDAE are taught to reconstruct input while promoting a language that is considerate and embraces diversity.

IV. VIDEO SUMMARIZATION TECHNIQUES USING DEEP LEARNING

The process of condensing lengthy videos into a concise representation that includes the essential elements of the original content is called video summarization. For various uses, including monitoring, information retrieval, and efficient

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

media consumption, understanding the relevance of something is critical. The field has seen remarkable strides in deep learning methods, leading to summarization that is both more accurate and sensitive to context. Here are a few popular deep learning-based methods for video summarization:

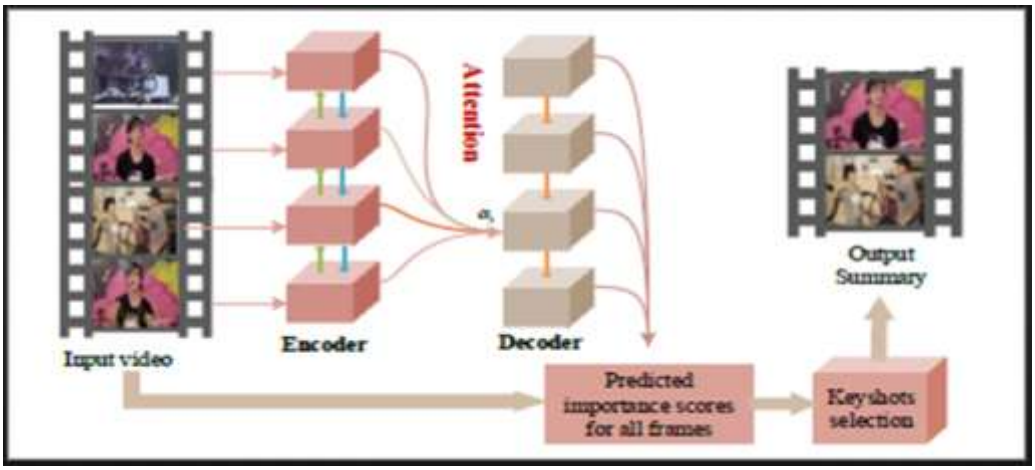

Fig. 4. Surprised video summarization using Deep learning

1. Temporal Segmentation and Clustering
Method: Deep learning models segment video into temporally coherent chunks or scenes and then cluster similar segments to create a summary.

Techniques:
Action Recognition: Models like 3D CNNs or I3D (Inflated 3D ConvNet) can recognize actions within segments to group similar scenes.
Temporal Convolutional Networks (TCNs): These networks can capture long-range temporal dependencies to segment video effectively.
Clustering Algorithms: Use algorithms like K-means or DBSCAN on feature embeddings from deep networks to cluster segments.

2. Video Captioning-Based Summarization
Method: Generate descriptive captions for video segments and use these captions to create summaries.
Techniques:
Sequence-to-Sequence Models: Employ LSTMs (Long Short-Term Memory networks) or GRUs (Gated Recurrent Units) with attention mechanisms to generate captions based on video features.
Transformer Models: Models like BERT or GPT can be adapted to generate coherent captions and summaries by understanding the video context.

3. Hierarchical Summarization
Method: Apply a hierarchical approach where videos are first summarized into smaller segments and then these summaries are combined to form a final summary.
Techniques:
Hierarchical Attention Networks: Use attention mechanisms at both the segment and video levels to focus on important parts of the video.
Two-Stream Networks: Combine spatial (frame-level) and temporal (sequence-level) features to capture high-level semantic information for summarization.

4. Keyframe Extraction
Method: Select keyframes that represent the most informative or representative moments in the video.
Techniques:
Recurrent Neural Networks (RNNs): Employ RNNs to analyze the sequence of frames and select those that capture significant changes or events.

5. Attention-Based Models
Method: Use attention mechanisms to focus on important parts of the video and generate a summary based on these focused segments.
Techniques:
Self-Attention Networks: Use self-attention to weigh the importance of different frames or segments within a video, as seen in Transformer-based architectures.
Temporal Attention: Apply attention mechanisms across time to identify crucial moments or events in the video sequence.

6. Graph-Based Methods
Method: Just use nodes to stand in for individual frames or segments and edges to show how they

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

relate to each other in time to make a movie diagram.
Techniques:
Graph Convolutional Networks (GCNs): Use GCNs to capture relationships between frames or segments and identify key segments for summarization.
Temporal Graph Networks: Apply graph-based approaches to capture temporal dynamics and summarize based on temporal relationships.

7. Reinforcement Learning
Method: Use reinforcement learning to automatically select the most informative segments of a video based on a reward mechanism.
Techniques:
Deep Q-Networks (DQN): Train an agent to maximize a reward function that measures the quality of the video summary.
Policy Gradient Methods: Employ policy gradients to optimize the selection of video segments or frames that contribute to an effective summary.

8. Multimodal Approaches
Method: Integrate multiple types of data (e.g., audio, text, visual) to create a comprehensive summary.
Techniques:
Multimodal Transformers: Combine features from different modalities (audio, video, text) to generate a more coherent and informative summary.
Cross-Modal Attention: Use attention mechanisms to align and integrate information from different modalities for improved summarization.

V. HATE SPEECH DETECTION IN DIFFERENT PLATFORMS

1. **Data Collection Methods:**
   - **Manual Annotation:** Gather a dataset of YouTube videos manually labeled for hate speech and non-hate speech. This involves watching videos and annotating segments or comments that contain hate speech based on predefined criteria.
   - **Crowdsourcing:** Use specialist annotation services or platforms like Amazon Mechanical Turk to improve the annotation process without sacrificing quality.
   - **Use of APIs and Tools:** Leverage YouTube Data API to access comments and metadata programmatically. Tools like YouTube Data Tools (YTDT) can help extract and categorize comments for analysis.
2. **Feature Extraction:**
   - **Audio Analysis:** There is a plethora of information retrieval methods and tools available for use when working with audio recordings. Some options to investigate include spectral features, Mel-frequency cepstral coefficients (MFCCs), and deep learning models designed for specific applications.
   - **Video Analysis:** Analyze visual content for hate symbols, gestures, or visual cues indicative of hate speech. Use computer vision techniques and pre-trained models for object detection and classification.

3. **Model Development:**
   - **Deep Learning Architectures:** Develop Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or Transformer-based models for multimodal analysis (audio, video, text) of YouTube content.
   - **Text Analysis:** comments or transcripts associated with videos. Use text classification models (e.g., BERT, LSTM) to detect hate speech in textual content.
4. **Evaluation and Monitoring:**
   - **Real-time Monitoring:** Implement systems for real-time detection and reporting of hate speech in YouTube videos. This may involve integrating models with YouTube's API to flag and moderate harmful content promptly.

Hate Speech Detection in News Articles:
1. **Data Collection Methods:**

News Aggregators and APIs: Use APIs from news aggregators (e.g., News API, GDELT) to collect articles from various sources. These APIs provide access to a broad range of news articles published online.
Manual Annotation: Manually label news articles as containing hate speech or not. This involves human judgment based on the content and context of the articles.
Web Scraping: Automate the collection of news articles from reputable sources or specific websites using web scraping techniques. Ensure compliance with terms of service and ethical considerations.

2. **Feature Extraction:**

Textual Analysis: Process news article text using NLP techniques to analysis, and topic modeling.
Contextual Analysis: Consider the context in which hate speech occurs, such as targeted groups, political discourse, or controversial topics. Develop models that capture these nuances.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

3. **Model Development:**

Machine Learning Models: Train supervised machine learning models (e.g., Naive Bayes, SVM, Random Forest) or deep learning models (e.g., BERT, CNNs) on labeled news articles to classify hate speech.

Topic Modeling: Discover trends in news articles and analyses hate speech using advanced approaches like Latent Dirichlet Allocation (LDA).

4. **Evaluation and Monitoring:**

Evaluation Metrics: Evaluate models using standard metrics appropriate for text classification tasks. Adjust evaluation criteria based on the specific goals and impact of hate speech in news media.

Ethical Considerations: Ensure ethical handling of data and model outputs, considering the potential impact of false positives and the need for transparency in automated content moderation.

Considerations for Both Platforms

- **Bias and Generalization:** Address bias in datasets and models to ensure fair and accurate detection across diverse content and user demographics.
- **Scalability:** Design scalable data collection and processing pipelines to handle large volumes of multimedia and textual data from YouTube and news sources.
- **Regulatory Compliance:** Adhere to legal and ethical guidelines regarding hate speech detection and moderation, respecting privacy and freedom of expression principles.

Example Study Design:

- **Research Question:** Can deep learning models effectively detect hate speech in YouTube video comments?
- **Hypothesis:** Deep learning models trained on annotated datasets can accurately identify hate speech in YouTube videos.
- **Methodology:** Conduct a comparative study using different deep learning architectures and evaluate their performance using standard metrics on a dataset of annotated YouTube video comments.
- **Data Collection:** Manually annotate a diverse dataset of YouTube video comments labeled for hate speech and non-hate speech.
- **Experimental Setup:** Implement and train CNNs, RNNs, and BERT-based models on the annotated dataset. Evaluate and compare their performance using accuracy, precision, recall, F1-score, and ROC-AUC metrics.
- **Ethical Considerations:** Address privacy concerns, mitigate biases in the dataset, and ensure compliance with ethical guidelines for hate speech detection.

VI. COMPARISON BETWEEN SPEECH DETECTION AND HATE SPEECH DETECTION

Detecting speech and detecting hate speech are two related but distinct tasks in natural language processing and content moderation. Here's a comparison between the two:

Speech Detection:

Refers to identifying and separating speech (spoken words) from non-speech segments in audio data.

Speech Recognition: Applied to voice-to-text systems, which include digital assistants such as Alexa and Siri. Automatic speech recognition (ASR), voice-to-text conversion, or speech recognition is a fascinating method. Recognition of spoken language, or the process of translating spoken sounds into a form that computers can understand.

Before further processing can begin, the audio stream undergoes a number of preprocessing steps to eliminate noise and enhance clarity. You may greatly enhance the signal quality by using particular methods like normalization, filtering, and noise reduction.

Using acoustic models, acoustic features derived from speech are linked to phonetic or sub-word units. In order to identify trends in audio inputs, these models use statistical.

Text created using state-of-the-art language models using audio input is shown below. They improve speech recognition accuracy by using their understanding of word sequences, grammatical rules, and contextual cues.

The decoded output is generated by combining acoustic and language models. It involves matching acoustic features with possible word sequences to produce the most likely transcription of the spoken input.

Audio Processing: Enables tasks such as audio indexing, transcription, and speech-to-text conversion.Audio processing refers to the manipulation and analysis of audio signals to extract meaningful information or enhance the quality of audio data. The primary objective of audio processing is to transform raw audio signals into a format that is more useful for specific applications.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 598-606 www.ijaem.net ISSN: 2395-5252

Challenges

Noise and Variability: Handling background noise, accents, and different speaking styles. Noise and variability are inherent characteristics of YouTube as a diverse and user-generated content platform. While noise can detract from user experience, strategies like improved algorithms, viewer tools, and creator best practices help mitigate its impact. Variability, on the other hand, contributes to the richness and diversity of content available on the platform, catering to a wide range of interests and preferences.Noise and variability are significant aspects of content on YouTube, impacting both content creators and viewers. Understanding these concepts can help in managing expectations, improving content quality, and enhancing user experience.

Speaker Independence: Recognizing speech from different speakers and adapting to individual voices.

Hate Speech Detection:

Involves identifying and categorizing text or speech that expresses or incites hatred, discrimination, or violence towards individuals or groups based on attributes like race, ethnicity, religion, gender, or sexual orientation.

Content Moderation: Used by online remove or abusive content.Content moderation on YouTube is a critical process aimed at ensuring the platform remains safe, respectful, and compliant with legal standards while balancing free expression. Given YouTube's vast and diverse user base, content moderation involves a combination of automated systems, human review, and community reporting.

Protecting users from harmful content such as violence, harassment, and hate speech.

Adhering to laws and regulations, including copyright, child protection laws, and regulations on hate speech.

Social Impact: Mitigates the spread of harmful ideologies and promotes a safer online environment.

Challenges:

- **Nuanced Language:** Identifying hate speech requires understanding context, sarcasm, and cultural references.
- **Bias and Fairness:** Addressing biases in training data and ensuring equitable detection across different groups.

Comparison:

- **Nature of Input Data:**
  - **Speech Detection:** Focuses on audio signals and spoken language.
  - **Hate Speech Detection:** Analyzes textual content, including written comments, articles, or transcripts.
- **Purpose:**
  - **Speech Detection:** Primarily facilitates interaction and communication through spoken language.
  - **Hate Speech Detection:** Aims to protect users from harmful and abusive content, promoting a respectful online community.
- **Complexity and Interpretability:**
  - **Speech Detection:** Relatively straightforward due to standardized speech patterns and acoustic features.
  - **Hate Speech Detection:** More complex due to the nuanced and context-dependent nature of language.
- **Ethical Considerations:**
  - Both tasks involve ethical considerations, such as privacy, freedom of expression, and avoiding algorithmic biases, but hate speech detection also focuses on societal impacts and fairness in content moderation.

While speech detection focuses on identifying spoken words in audio signals, hate speech detection aims to identify harmful language in textual content, addressing different challenges and societal concerns related to online discourse and content moderation. Each task requires tailored approaches in data collection, feature engineering, model development, and evaluation metrics to achieve effective detection and moderation capabilities.

CONCLUSION

The Hybrid Deep Ensemble (HDE) is a novel algorithm designed to categorize YouTube videos into various categories. It uses a cascade architecture of base learners, where each model improves its performance by analyzing the predictions of the previous model. This approach allows for targeted improvements to existing labels, enhancing overall performance. The research used a cascade structure to construct machine learning models for filtering YouTube material, similar to that of a data scientist.The effectiveness of the HDE was compared to individual machine learning and deep learning models, and the study used a small dataset. Standard models, such as multilayer perceptron, deep neural network, and Keras deep learning classifier, showed similar performance to the HDE. The HDE demonstrated remarkable versatility and exceptional classification accuracy, surpassing expectations despite minimal training data.The system's effectiveness exceeded

expectations, and the computational complexity was precise. When applied to a dataset with limited data, the HDE demonstrated outstanding classification accuracy. This application could be beneficial for speech-language pathologists who need YouTube video samples for assessment purposes. The frequency distribution of each YouTube video could provide valuable data on participants' fluency scores. This content is a valuable resource for those seeking to enhance their YouTube watching experience.

REFERENCES

[1]. Banks, James. "Regulating hate speech online." In: International Review of Law, Computers Technology 24.3, pp. 233-239,2010.

[2]. Caselli, T., Basile, V., Mitrović, J., Granitzer, M., 2020. Hatebert: Retraining bert for abusive language detection in english. arXiv preprint arXiv:2010.12472.

[3]. Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., Bhamidipati, N., 2015. Hate speech detection with comment embeddings, in: Proceedings of the 24th international conference on world wide web, pp. 29–30.

[4]. Founta, P. et al.," "Large scale crowdsourcing and characterization of twitter abusive behavior." Proceedings of the international AAAI conference on web and social media. Vol. 12. No. 1. 2018.

[5]. Ishmam, A.M., Sharmin, S., 2019. Hateful speech detection in public facebook pages for the bengali language, in: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), IEEE. pp. 555–560.

[6]. Liu, Y, Shriberg, E, Stolcke, A & Harper, M 2005, 'Comparing HMM, maximum entropy, and conditional random fields for disfluency detection', Proceedings of ninth European Conference on Speech Communication and Technology, pp. 3313-3316.

[7]. Mondal, M., Silva, L. A., & Benevenuto, F. (2017, July). A measurement study of hate speech in social media. In Proceedings of the 28th ACM conference on hypertext and social media (pp. 85-94).

[8]. Prasadu Peddi, & Dr. Akash Saxena. (2016). Studying data mining tools and techniques for predicting student performance. International Journal Of Advance Research And Innovative Ideas In Education, 2(2), 1959-1967.

[9]. Yin, W., Kann, K., Yu, M., Schütze, H., 2017. Comparative study of cnn and rnn for natural language processing. arXiv preprint arXiv:1702.01923.

[10]. Wang, S., Liu, J., Ouyang, X., Sun, Y., 2020. Galileo at semeval-2020 task 12: Multi-lingual learning for offensive language identification using pre-trained language models. arXiv preprint arXiv:2010.03542.

[11]. Y. Zhou, Y. Yang, H. Liu, X. Liu, and N. Savage, "Deep learning based fusion approach for hate speech detection," IEEE Access, vol. 8, pp. 128 923–128 929, 2020.

[12]. Zhu, JY, Park, T, Isola, P & Efros, AA 2017, 'Unpaired image-to-image translation using cycle-consistent adversarial networks', Proceedings of the IEEE international conference on computer vision, pp. 2223-2232.