Edit report draft (Latex) through this link:

[*https://www.overleaf.com/3376294213qyvxgpcndgyn*](https://www.overleaf.com/3376294213qyvxgpcndgyn)

**Abstract**

Content creators grapple with the challenge of predicting if their time and investments will translate into increased viewership and audience growth, a task made more complex by the hidden algorithms and unpredictable audience interaction of social media platforms. This research's objective is to architect a model that predicts video success, effectively indicating a video's potential virality. By employing advanced convolution techniques for video encoding, leveraging strides in natural language processing models, we’re pushing boundaries in deep video content analysis. We construct a powerful multimodal ensemble model for general video content regression, capable of comprehending human-generated content and accurately predicting its nonlinear relationship elements to determine success. Our preliminary results demonstrate the model's effectiveness in predicting video virality, showcasing the potential of our innovative approach.

**1. Introduction**

**2. Approach**

2.1 Data collection and preparation

2.2 Model

2.2.1 Visual Embedding

2.2.2 Audio Embedding

2.2.3 Transformer-based Regression

2.3 Loss Function

2.4 Implementation

**3. Experiments and Results**

3.1 Experiment Setup

3.2 Results

**4. Conclusion and Future Improvements**

**5. Work Division**

Contributions of each group member can be found in Table below.

|  |  |  |
| --- | --- | --- |
| Student Name | Contributed Aspects | Details |
|  |  |  |

Table Contributions of team members

**References**

Figure Illustration of model

Figure Training loss, validation…