Personalizing Music Video Recommendations with Emotional Intelligence

Lionel

Background

- Music recommendation is a critical area where emotions play a pivotal role, as it's common for individuals to choose music that **resonates with their current mood** (Juslin, Sloboda et al. 2001)
- However, existing social media video recommendation methods, such as those on YouTube, primarily rely on genre features. This approach may not fully cater to the emotional needs of users, as it overlooks the emotional connection that users often seek in music.
- Significant opportunity to enhance these systems by incorporating emotion-based recommendations, leading to a more personalized and emotionally resonant user experience.

Problem Statement

The objective is to develop an **emotion-centric** music video recommendation system. This system aims to identify a YouTube user's emotional state from their comments and recommend music videos that either **align with their current emotional state** or assist them in **overcoming negative emotions**.

- The system will prioritize user well-being by suggesting songs that harmonize with their current mood. It will evaluate users' emotional needs based on their comments and recommend the most "beneficial" music videos instead of the most "matched" ones.

Why this matters

By focusing on the emotional needs of users, this recommendation system goes beyond genre-based recommendations, offering a more nuanced and empathetic approach to music video suggestions. This could potentially revolutionize the way users interact with music on social media platforms.

Who it Matters to?

	Melody	Harmony
who	A music enthusiast who uses YouTube as a primary source for discovering new music.	A member of the YouTube Product Team responsible for improving user engagement and satisfaction.
pains	Often feels overwhelmed by the vast number of music video recommendations that don't align with her current mood or emotional state.	Struggles with the challenge of increasing user engagement and personalizing the user experience.
goal	Wants a more personalized music recommendation experience that understands and caters to her emotional needs.	Wants to implement a system that can accurately gauge users' emotional states and recommend music videos accordingly, thereby enhancing user engagement and satisfaction.

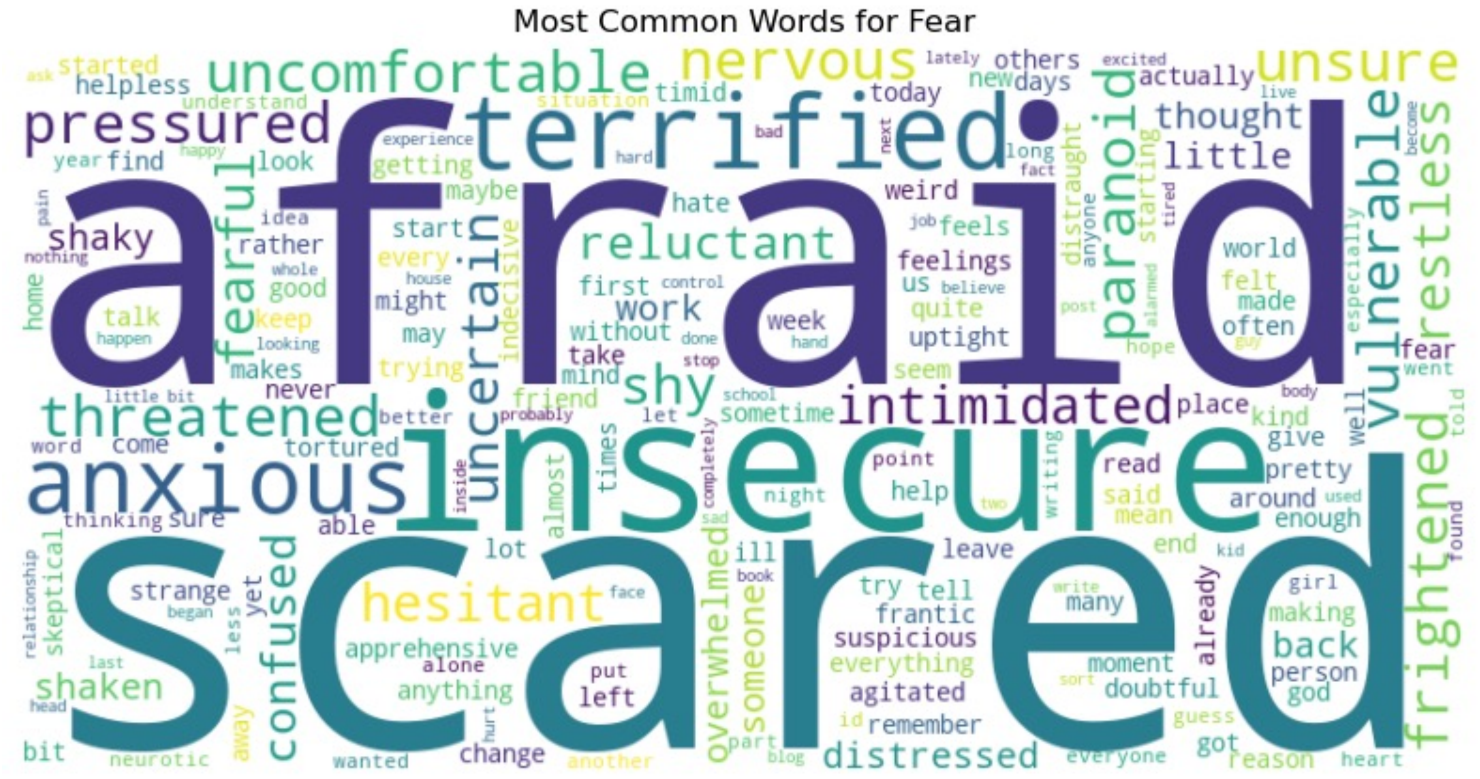
Modeling

- Emotion Classification Model
- Rule-Based Music Video Recommender System

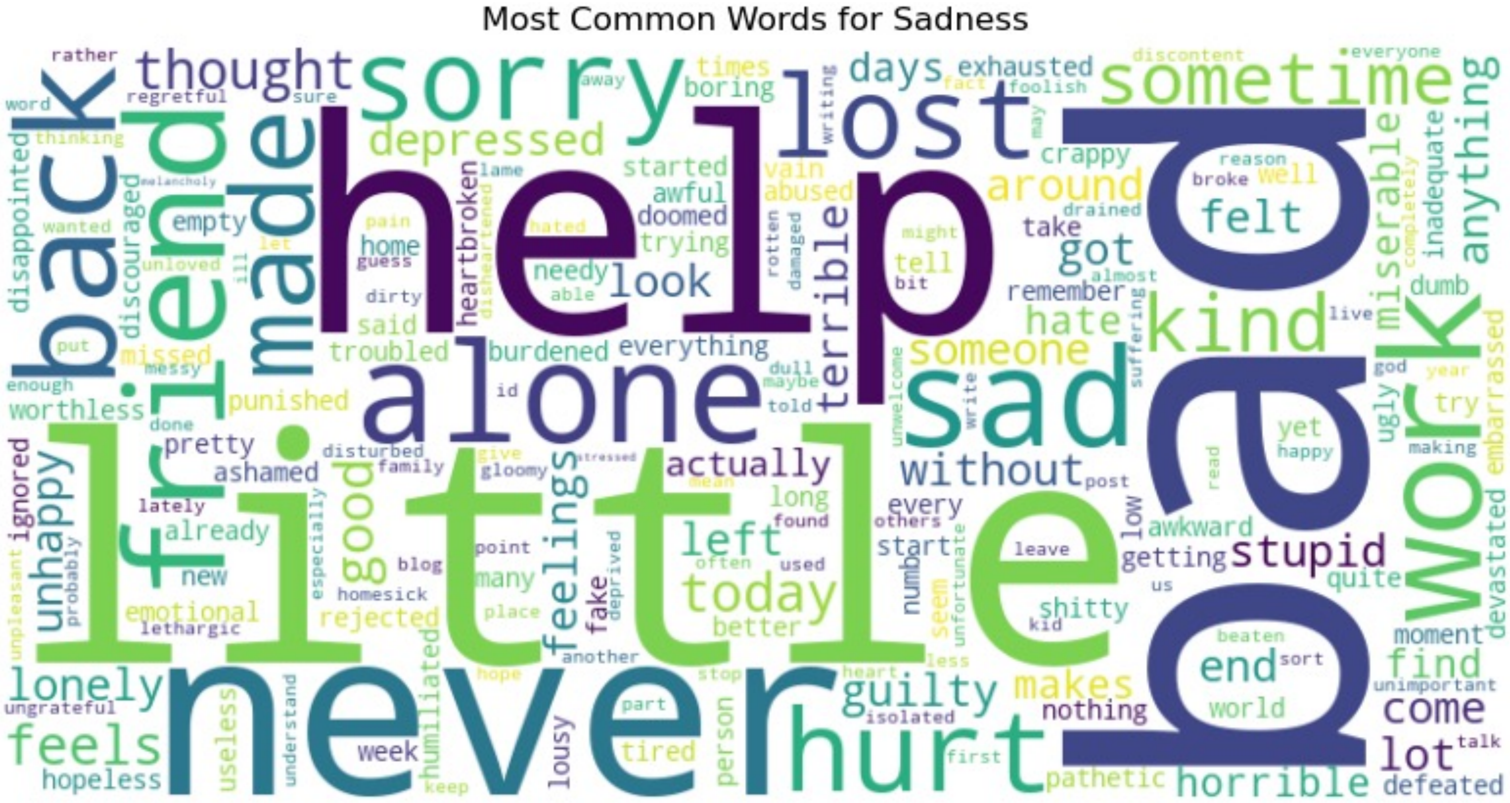
Data - Source

- Emotions Dataset from Kaggle
- This dataset contains English Twitter messages annotated with six fundamental emotions: Sadness, Joy, Love, Anger, Fear & Surprise

Data – Fear Most Common Words

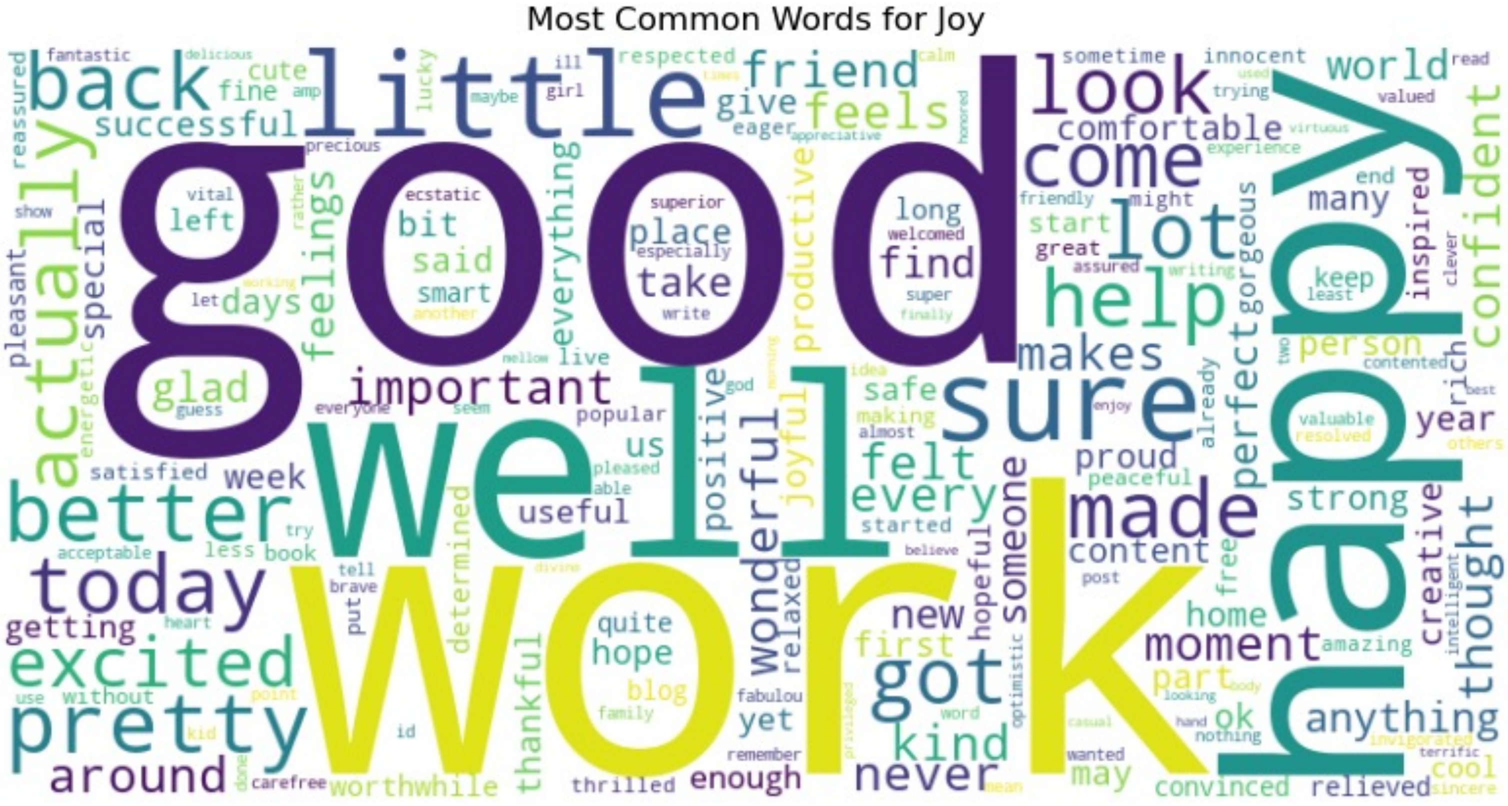


Data – Sadness Most Common Words

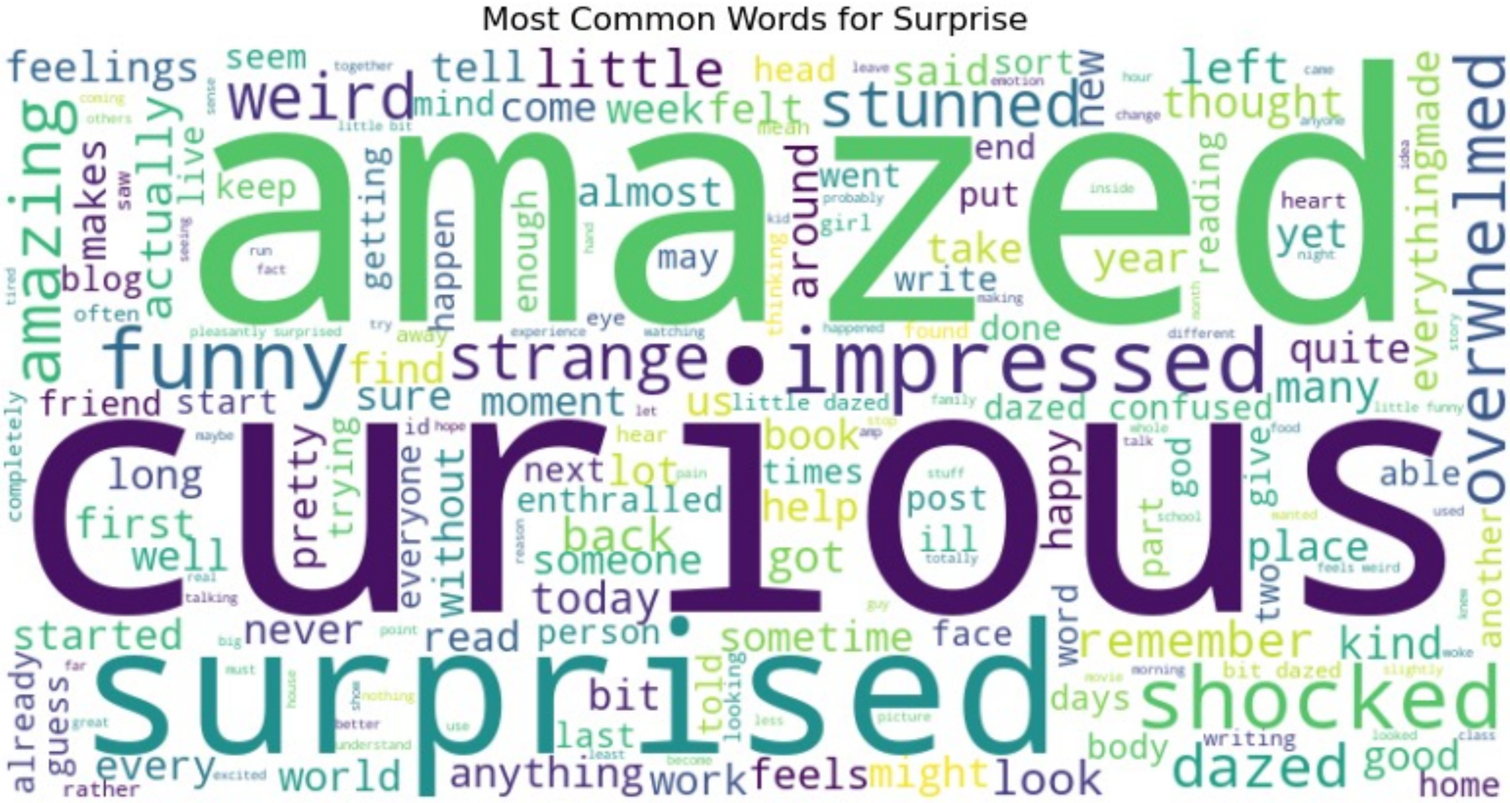


Data – Love Most Common Words

Data – Joy Most Common Words



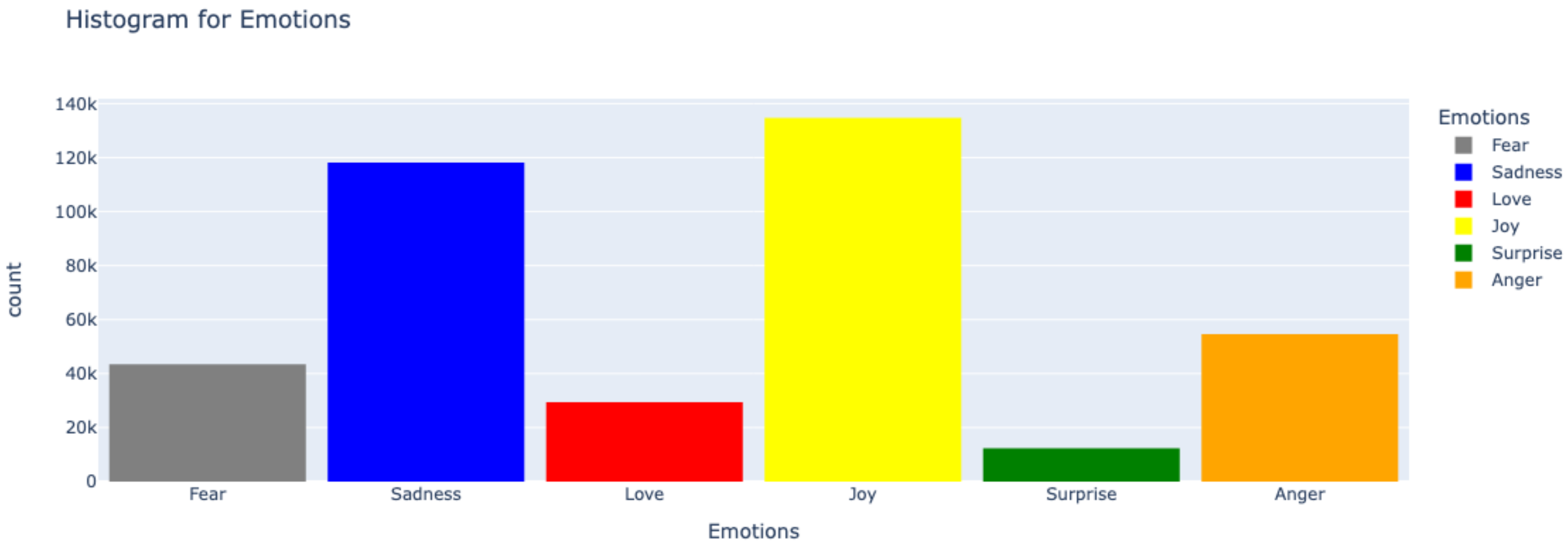
Data – Surprise Most Common Words



Data – Anger Most Common Words

Most Common Words for Anger

Distribution of Emotion Classes



Approach to Addressing Class Imbalance

- *Surprise* emotion is significantly less common in the dataset than the other emotions
  - To balance dataset by removing all rows with 'Surprise' class
  - Surprise is a transient emotion
  - Surprise can have any valence

- Undersampling (oversampling) of the most (least) numerous class to attain a count equivalent to the *Fear* class

Modeling

- Fine tuning DistilBERT
    - Can capture the context of words in a sentence
- Multinomial Naive Bayes
- Logistic Regression
- XGBoost
- RNN-LSTM with GloVe Word Embedding

Evaluation Metrics of Interest

Metric	Description / Rationale
Accuracy	% of predictions that are correct
F1-Score	Composite measure that considers both (a) "precision" and (b) "recall" (a) "precision": % of correct predictions for a particular emotion out of all predictions for that emotion *Relevant as a we want to keep the number of incorrect predictions low* (b) "recall": % of correct predictions for a particular emotion out of all actual instances of that emotion *Relevant as we wish to identify as many instances of each emotion as possible*
Efficiency	Computation time

Modeling Results

	Training accuracy	Testing accuracy	Precision	Recall	F1 score	Computation time
Multinomial Naive Bayes	0.95	0.93	0.90	0.94	0.91	1.5 x
Logistic Regression	0.97	0.94	0.92	0.95	0.93	1 x
XGBoost	0.95	0.93	0.90	0.94	0.92	2 x
DistilBERT	0.44	0.57	0.62	0.57	0.58	30 x
RNN-LSTM with GloVe Word Embedding	1.00	0.97	0.95	0.97	0.96	2.7 x

Finalised Model

	Training accuracy	Testing accuracy	Precision	Recall	F1 score	Computation time
RNN-LSTM with GloVe Word Embedding	1.00	0.97	0.95	0.97	0.96	2.7 x

- Combines the strengths of RNNs and GloVe embeddings
- RNNs are capable of capturing the sequential information present in the text data
- GloVe can capture semantic relationships between words

Rule-Based Music Video Recommender System

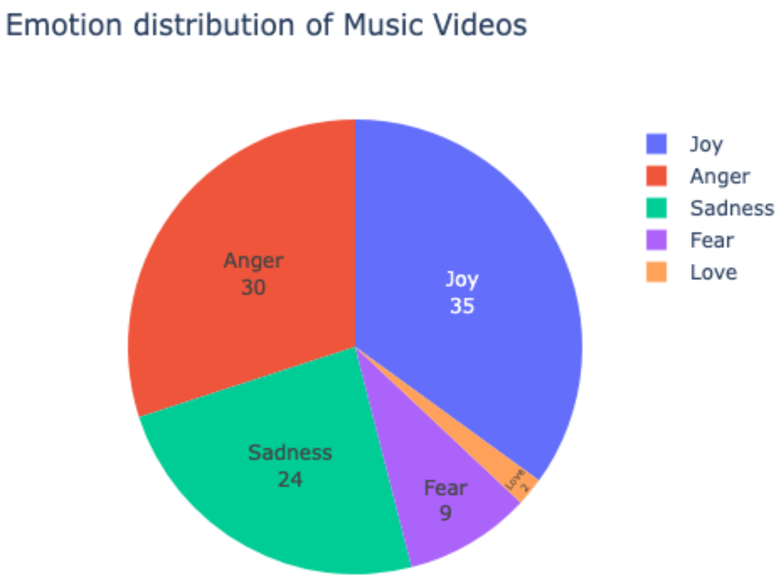
- Transcripts Scraped from 100 Youtube Music videos (popular music videos from 2020s)
- Scraped Comments from each of the 100 Music Videos

Using Trained Emotion Classification Model to predict emotions

Cleaned Transcripts → Saved Tokenizer → Saved LSTM Model → Predicted Emotion

Cleaned Comments → Saved Tokenizer → Saved LSTM Model → Predicted Emotion

Distribution of Music Video Emotion Classes



Music Video Transcripts and predicted emotions

Video_id	Transcript	Probability -Sadness	Probability- Joy	Probability- Love	Probability - Anger	Probability - Fear	Predicted Emotion
G7KNmW9a75Y	Lyrics 1	0.19	0.60	0.03	0.16	0.02	Joy
Oa_RSwwpPaA	Lyrics 2	0.33	0.34	0.01	0.06	0.26	Joy
8UMnAIaBofo	Lyrics 3	0.30	0.14	0.11	0.36	0.09	Anger
Iq8h3GEe22o	Lyrics 4	0.41	0.02	0.01	0.55	0.01	Anger
H5v3kku4y6Q	Lyrics 5	0.34	0.30	0.05	0.09	0.22	Sadness

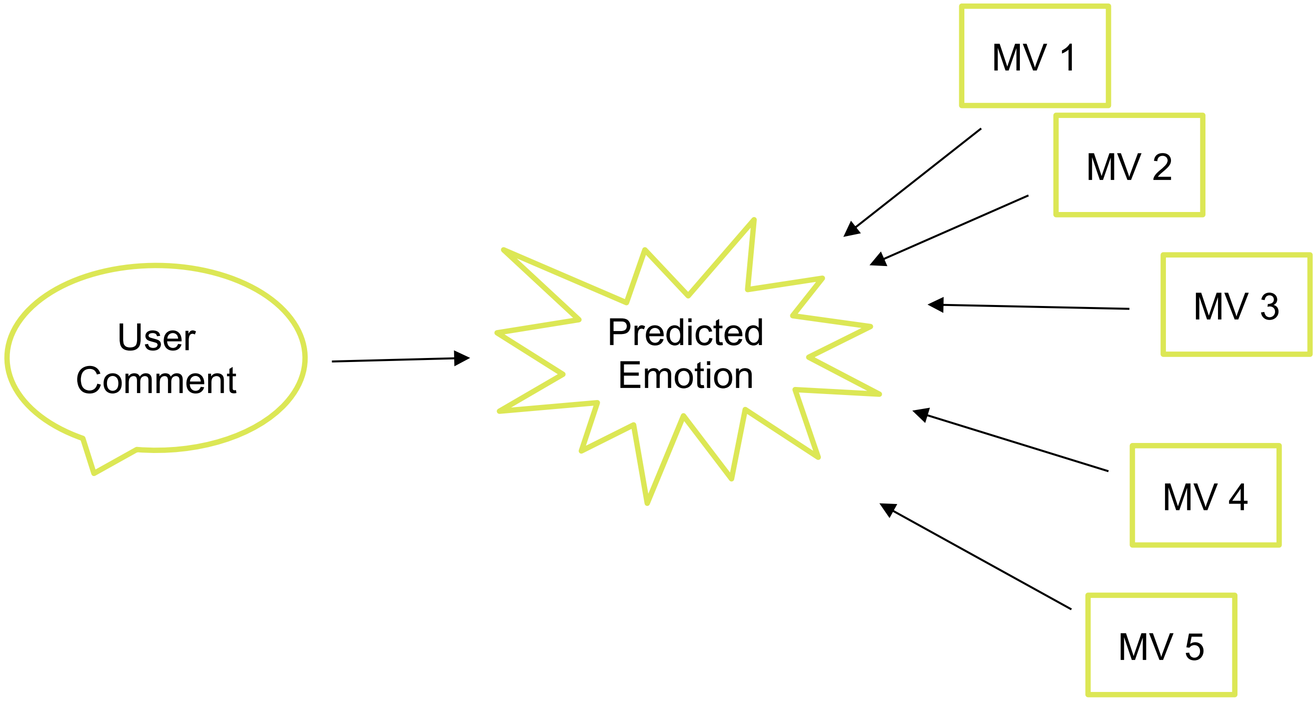
Music Video comments and predicted emotions

author	comment	Probability -Sadness	Probability-Joy	Probability-Love	Probability - Anger	Probability - Fear	Predicted Emotion
@Royalchess1	Comment 1	0.54	0.01	0.03	0.13	0.29	Sadness
@mikerooney7600	Comment 2	0.02	0.95	0.01	0.01	0.01	Joy
@Ganesh-zs1iv	Comment 3	0.13	0.01	0.01	0.85	0.02	Anger
@helensaunders5085	Comment 4	0.07	0.05	0.67	0.08	0.13	Love
@feully	Comment 5	0.25	0.05	0.02	0.12	0.56	Fear

Emotion Based Recommendation System

- This system will analyse users' mental states (based on comments) and identify their emotional needs. Instead of recommending the most "matched" music, we will suggest the most "helpful" music videos. The most helpful music video should meet two criteria:
    - it aligns with the user's mental state, meaning it satisfies the user's emotional needs
    - If the user's overall mood is negative, the recommended music video should have the potential to uplift the user's mood. This implies that the music should be slightly more positive than the user's current mood. However, to prevent causing emotional discomfort or resistance to the recommendation, the degree of positivity should be delicately balanced.

Rule Based Recommendation System



Emotions are categorised into 2 groups

User Comment
Positive Emotion

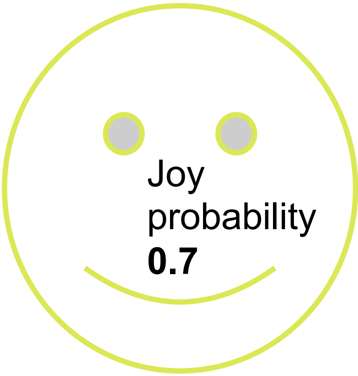
Joy, Love

User Comment
Negative Emotion

Sadness, Anger, Fear

User’s Emotion is positive (Joy or Love)

- Look for music videos that evoke a similar emotion



User's emotion is negative (Sadness or Anger or Fear)

- look for music videos that can help uplift their mood.
- Only consider videos whose sum of of negative emotion probabilities is at least 0.1 less than that of the user's comment.

User comment $\begin{bmatrix} \text{Prob Sadness} \\ \text{Prob Anger} \\ \text{Prob Fear} \end{bmatrix}$ **Music Video** $\begin{bmatrix} \text{Prob Sadness} \\ \text{Prob Anger} \\ \text{Prob Fear} \end{bmatrix}$ $=$ **Cosine Similarity**

Demo time

Limitations

- The classification model was trained on a labeled dataset of tweets from Kaggle. Might not be representative of the language used in YouTube comments or music video lyrics.
- Complexity of Emotions
    - Emotions can overlap and it can be challenging to discern emotions from text.
- Limited Scope of Analysis
    - The system only analyzes the transcripts of the music videos, which might not fully capture the emotional content of the songs.

Future Work

- Consider increasing the data pool to include labeled comments from other social media platforms
- Explore using unsupervised learning to cluster the emotions of the text
- Incorporate Audio and Visual Analysis

Conclusion

- Developed a classification model to predict emotions derived from text data.
- Applied this model to discern emotions associated with users and music videos.
- Utilized the predicted emotions to construct the recommendation system.
- Existing social media video recommendation methods, such as those on YouTube, primarily rely on genre features. This approach may not fully cater to the emotional needs of users, as it overlooks the emotional connection that users often seek in music.
- Offers and alternative by recommending music videos that align with the user's emotional state and uplift their mood if they have negative feelings

References

Background

1. https://psycnet.apa.org/record/2001-05534-001
2. https://www.utu.fi/en/news/press-release/music-can-evoke-strong-emotions-the-same-melody-can-lead-to-shivers-down-the-spine-or-feelings-of-joy
3. https://ieeexplore.ieee.org/document/9256342
4. https://link.springer.com/chapter/10.1007/978-3-030-35288-2_20
5. https://www.sciencedirect.com/science/article/abs/pii/S030645732200357O