VIRAL MELODIES: EXPLORING THE FACTORS INFLUENCING MUSIC VIRALITY IN TIKTOK ENGAGEMENT

Research and Methodology in Computer Science

Team Members :

- 2540118914 - Evelyn Zefanya Rahardjo
- 2501980690 - Jasmine Mutia Alifa
- 2501961022 - Sharon Zefanya Setiawan

INTRODUCTION

Background

TikTok is a popular social media platform with billions of active users. Music plays a central role in the user experience, helping to create emotion, increase appeal, and connect content with the audience. Knowing what factors influence virality can increase user engagement and is a key factor in content success.

Research Purposes

1. Investigate what factors contribute to the popularity and virality of content on TikTok (beat, share, artist).
2. Find out what variables have a strong correlation.
3. Understand Virality Mechanisms on TikTok

Methodology

1. Use of Kaggle Data: Analyzing datasets from Kaggle to understand the impact of music on engagement.
2. Descriptive Analysis: Uses descriptive statistical techniques such as bar graphs, pie charts, means, medians, and modes to interpret data.

Contribution

1. This research provides a deeper understanding of what makes music content go viral on TikTok. It expands our knowledge of social media dynamics and virality, especially in a musical context.
2. Provides Guidance for Content Strategy: By identifying the key factors that contribute to virality, this paper provides valuable guidance for content creation and distribution strategies on social media.

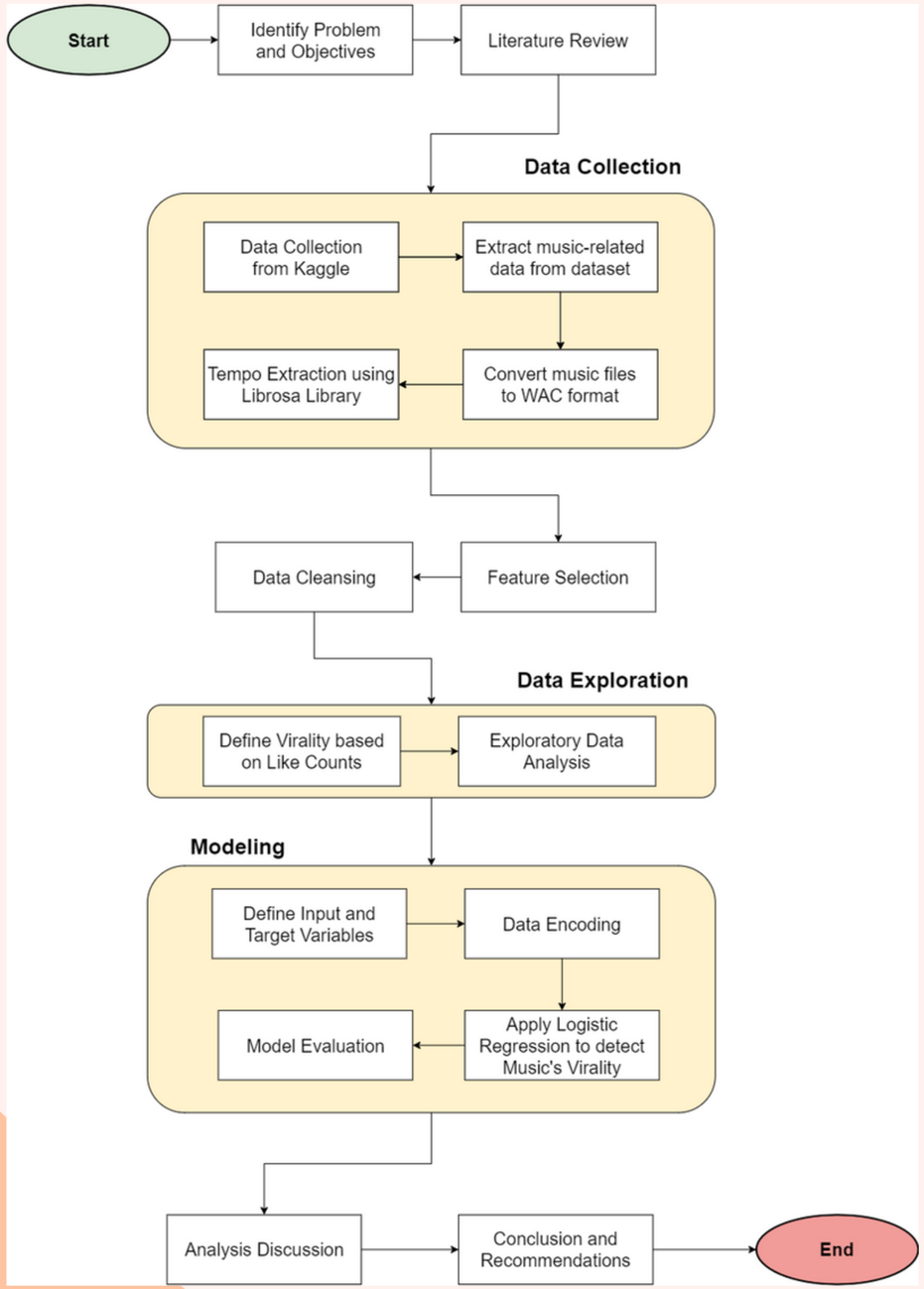
RELATED WORKS

1. Research conducted by the Institute of Musicology SASA, Belgrade, Serbia discusses the influence of music in the context of using TikTok, highlighting how TikTok has influenced the process of making music, listening to music, and promoting music.

2. Another research conducted by Han Yang focuses on the potential for innovation in music marketing via the TikTok platform. This study shows that music can be used as a medium to attract user's attention and can influence popularity on TikTok. This means that music can indeed influence user interest and increase engagement.

3. A study conducted by Terrence Cook found that music preferences play an important role in how people express their emotions. The conclusion is that music plays an important role in emotions.

METHODOLOGY



Research Flow Diagram

METHODOLOGY

DATA COLLECTION

TikTok Trending Videos Dataset

https://www.kaggle.com/datasets/erikvdven/tiktok-trending-december-2020/data

Type of Dataset

- Audio dataset from Spotify and Apple Music in CSV format.
- Videos from TikTok in MP4 format
- Collection of trending videos in JSON format.

Json Data : Consisting of 5693 records with 17 characteristics.

Variable	Description
URL	Link to the audio files
webVideoUrl	Link to the Tiktok videos
Filename	Music in format WAV
artist	Artist of the audio tracks
title	Title of the audio tracks
likeCount	Amount of likes
shareCount	How many times the video has been shared
playCount	How many times the video has been watched
commentCount	Amount of comments
bpm	A measure of speed or tempo in music
beat	Beats or pulses in a musical composition

Steps:

- Convert audio files from URL to WAV format using AudioSegment.from_file.
- Created a beat_track function to calculate beat values. The beat_track function detects the tempo (BPM) of a WAV audio file to determine its beat category. BPM calculation via librosa.beat.beat_track.
- Create new dataset. The variables in the new dataset are the variables that have the strongest correlation with the virality of a video.

METHODOLOGY

DATA PREPROCESSING

In this data analysis project, we use Python and libraries like pandas, numpy, re, and unidecode for data pre-processing. Data pre-processing is carried out to prepare the data before carrying out analysis. The goal is to address missing data, normalize, and handle unusual data.

Analysis Step	Description
Text Data Cleaning	Uses unidecode to replace accent characters with non-accent characters, as well as remove special characters, foreign symbols, and extra spaces.
Convert Text Data to Numeric	Convert social media data such as "likes", "comments", and "shares" from text to numeric format for more accurate analysis.
Data Type Refinement	Converts numeric data from floating-point to integer format for more precise representation and maintaining data integrity.
Duplicate Entry Handling	Identifies and eliminates duplicate entries, especially in the 'webVideoUrl' category, to reduce redundancy and increase the uniqueness and completeness of the dataset.

METHODOLOGY

DATA EXPLORATION

The purpose of doing EDA is to understand data distribution, create graphic visualizations, understand descriptive statistics, investigate relationships and trends, and identify features.

Steps

- Classification into viral or non-viral categories is based on 'likeCount', which results in the creation of a 'virality' column. The way it is created uses the detect_virality function, which categorizes content as 'Viral' or 'Nonviral' based on 'likeCount', with 302 videos falling into the 'Viral' category.
- Create a visualization in the form of a bar graph that displays the distribution of 'beat' categories in viral and non-viral content.
- Create visualizations in the form of pie charts to compare tempo distribution in viral and non-viral videos.
- Create a correlation matrix to determine the strong relationship between variables.
- Creation of a new column namely 'ShareGroup' to categorize values from the 'shareCount' column into 'Low', 'Medium', or 'High'.

METHODOLOGY

DETERMINING INPUT AND OUTPUT

In logistic regression, determining the input (independent variable) and output (dependent variable) is an important step to understand the cause-and-effect relationship between two or more quantitative variables. Inputs are factors that we believe influence the variables we observe or predict (output). The following is a table for determining input and output:

Variable	Description	Input / Output
artist	Artist of the audio tracks	Input → X
title	Title of the audio tracks	Input → X
shareCount	How many times the video has been shared	Input → X
virality	explanation of viral or non-viral video	Output (Target Variable) → y

METHODOLOGY

LOGISTIC REGRESSION

Logistic regression, a statistical method, is employed in music analysis to discern the factors influencing song virality. With independent variables including 'beat', 'artist', and 'shareGroup', and the dependent variable is 'virality'. Using Python, the data was processed by converting independent variables to categorical data and converting dependent variable to binary. After removing irrelevant columns, a design matrix was created using the patsy library, dividing the data set into dependent and independent variables for analysis.

METHODOLOGY

EVALUATION

In the post-modeling phase, the evaluation process involves generating a model summary using the summary function. This summary includes essential evaluation metrics such as Df Residuals, Df Model, Pseudo R-squ, Log-Likelihood, LL-Null, LLR p-value, coefficient, standard error, t, P>|t|, upper bound, and lower bound. These metrics collectively offer insights into model fit, the significance of independent variables, and the model's capacity to elucidate data variability.

RESULT AND DISCUSSIONS

A. VISUALIZATION

The purpose of employing visualizations is to gain a comprehensive understanding of the underlying patterns and relationships within the dataset.

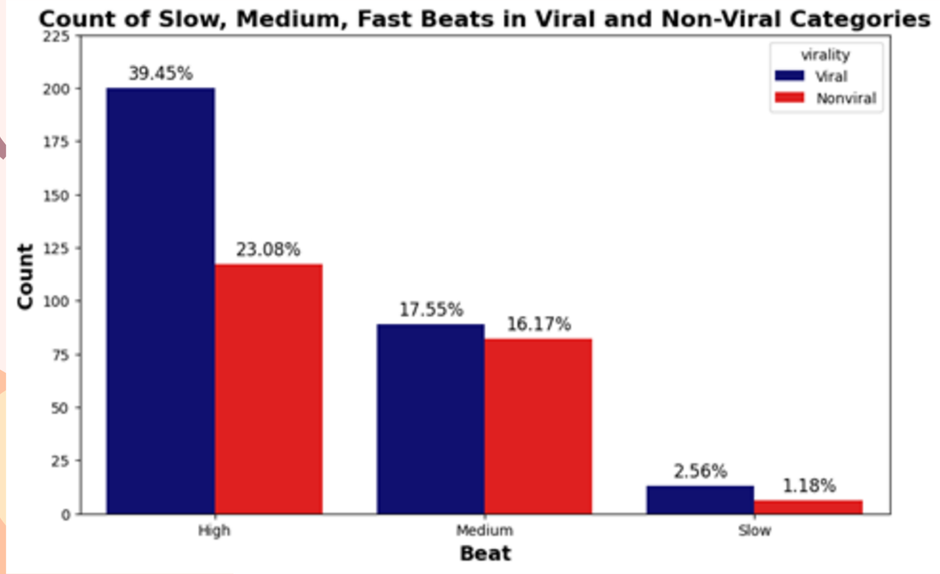


FIGURE 1

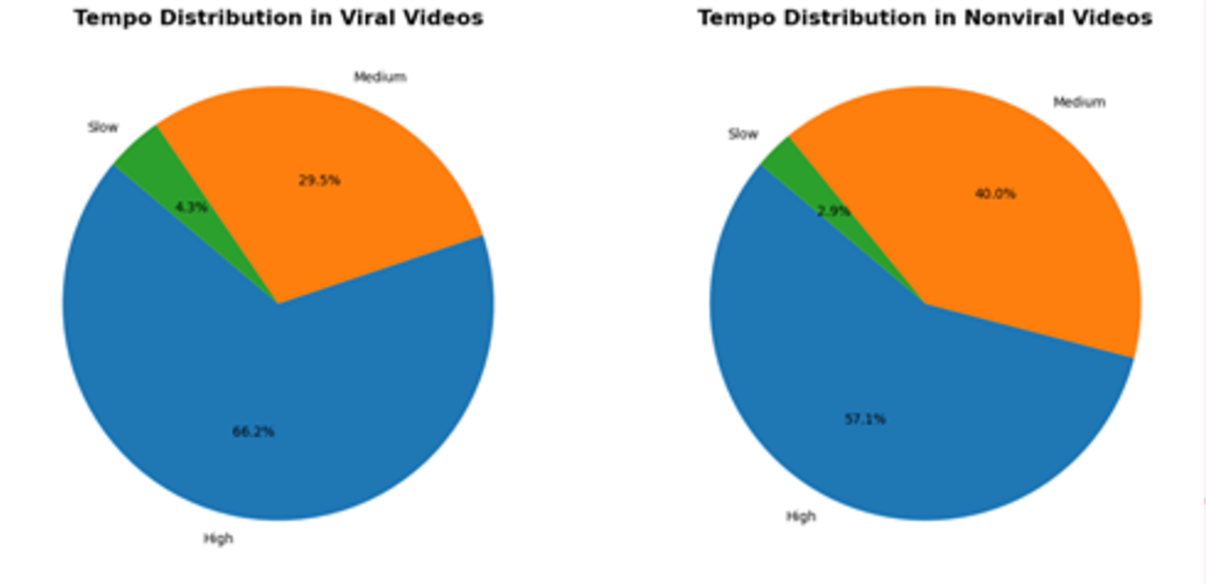


FIGURE 2

RESULT AND DISCUSSIONS

A. VISUALIZATION

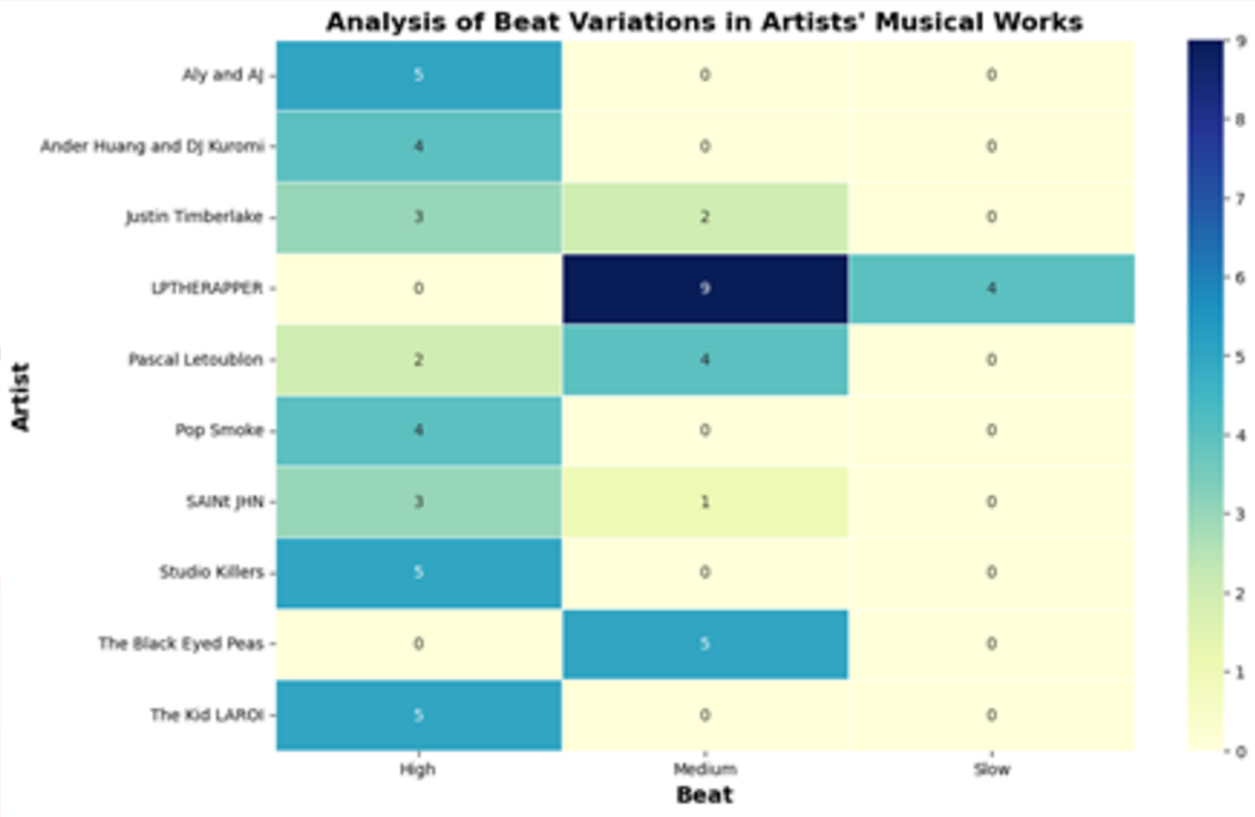


FIGURE 3

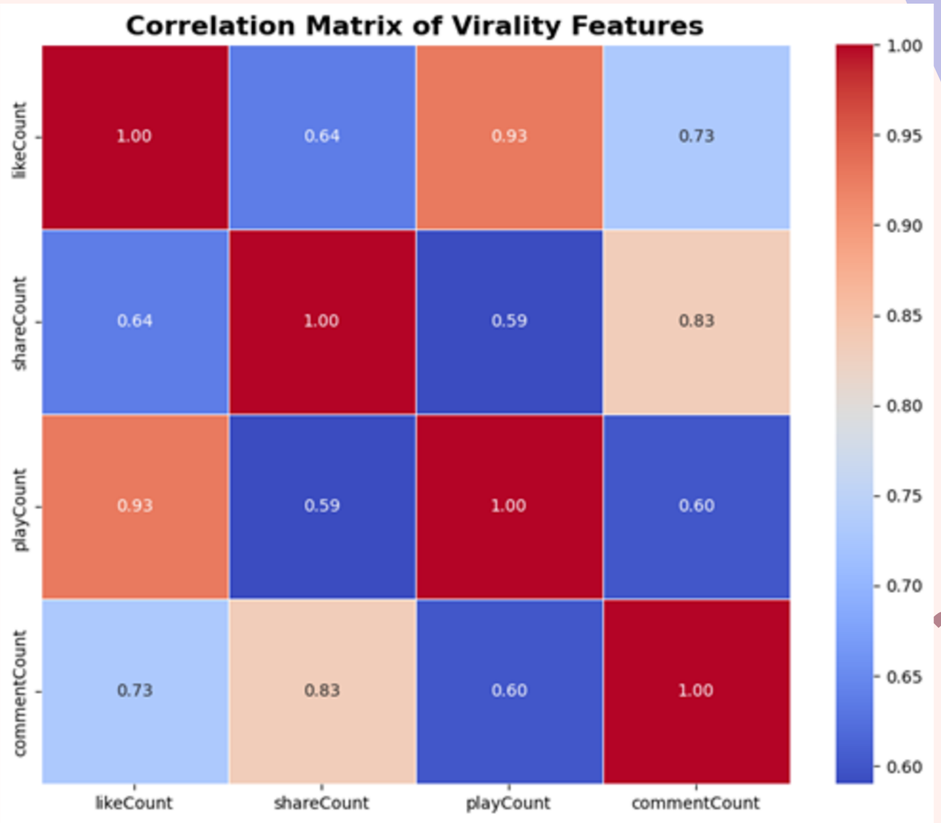


FIGURE 4

RESULT AND DISCUSSIONS

B. LOGISTIC REGRESSION SUMMARY MODEL

Logistic Regression Results						
Df Residuals	503					
Df Model	3					
Pseudo R-squ.	0.3123					
Log-Likelihood	-235.26					
LL-Null	-342.09					
LLR p-value	4.736e-46					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.7279	0.244	-2.984	0.003	-1206	-0.250
artist	0.0008	0.001	0.750	0.453	-0.001	0.003
beat	-0.3214	0.205	-1.570	0.117	-0.723	0.080
shareGroup	2.4107	0.235	10.276	0.000	1.951	2.870

CONCLUSION

This study investigates the virality dynamics of music content on TikTok, emphasizing beat, artist, and social shares, utilizing the librosa library for tempo extraction and logistic regression for analysis. The findings indicate that 'shareGroup' significantly influences virality, while 'beat' and 'artist' show no significant impact, contributing to a moderately good fit of the model, explaining approximately 31.23% of the variability in virality. Future research suggestions include a deeper exploration of TikTok's social dynamics, particularly in understanding video-sharing mechanisms among users with similar music preferences for enhanced content strategies.

THANK YOU