

# Genre Prediction From Lyrics

By  
Teja Kalvakolanu  
Bao Nguyen  
Nagasai Chandra

# INTRODUCTION

Prediction of Genre from Lyrics

Dataset from kaggle 380,000+ lyrics from Metro Lyrics

Reasoning for prediction

Recommend songs with similar lyrics

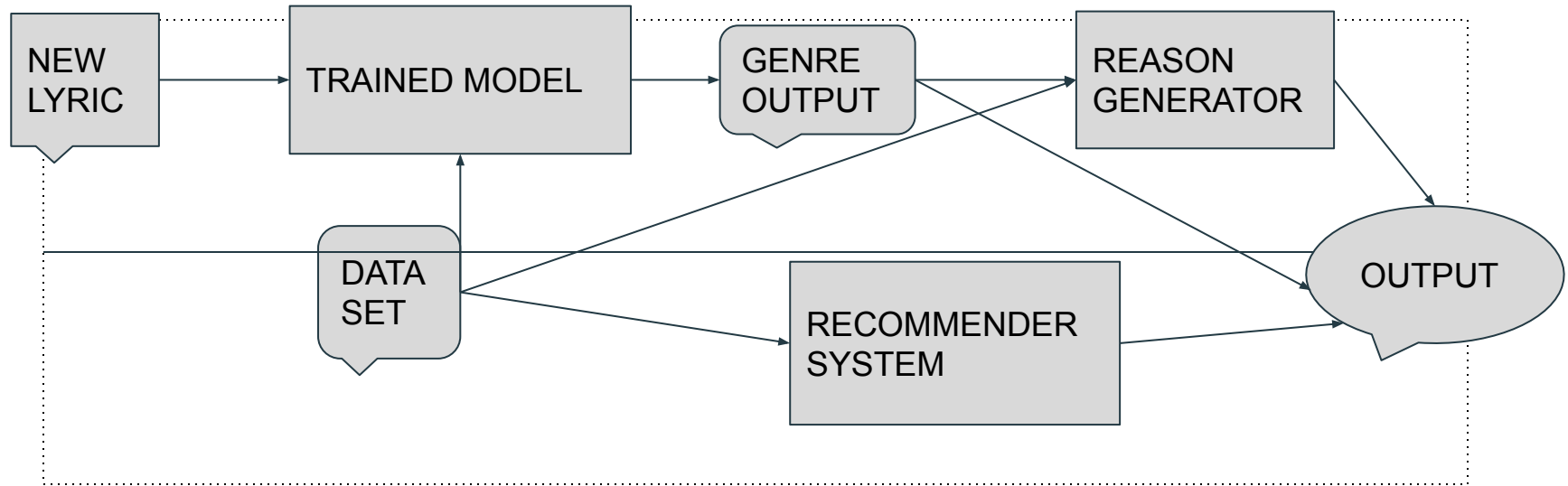
Open problem for research

# BACKGROUND/RELATED WORK

In general audio features is common used to classify a song while using lyrics is a challenging.

Previously researchers didn't provide supporting reasons

# SYSTEM DIAGRAM



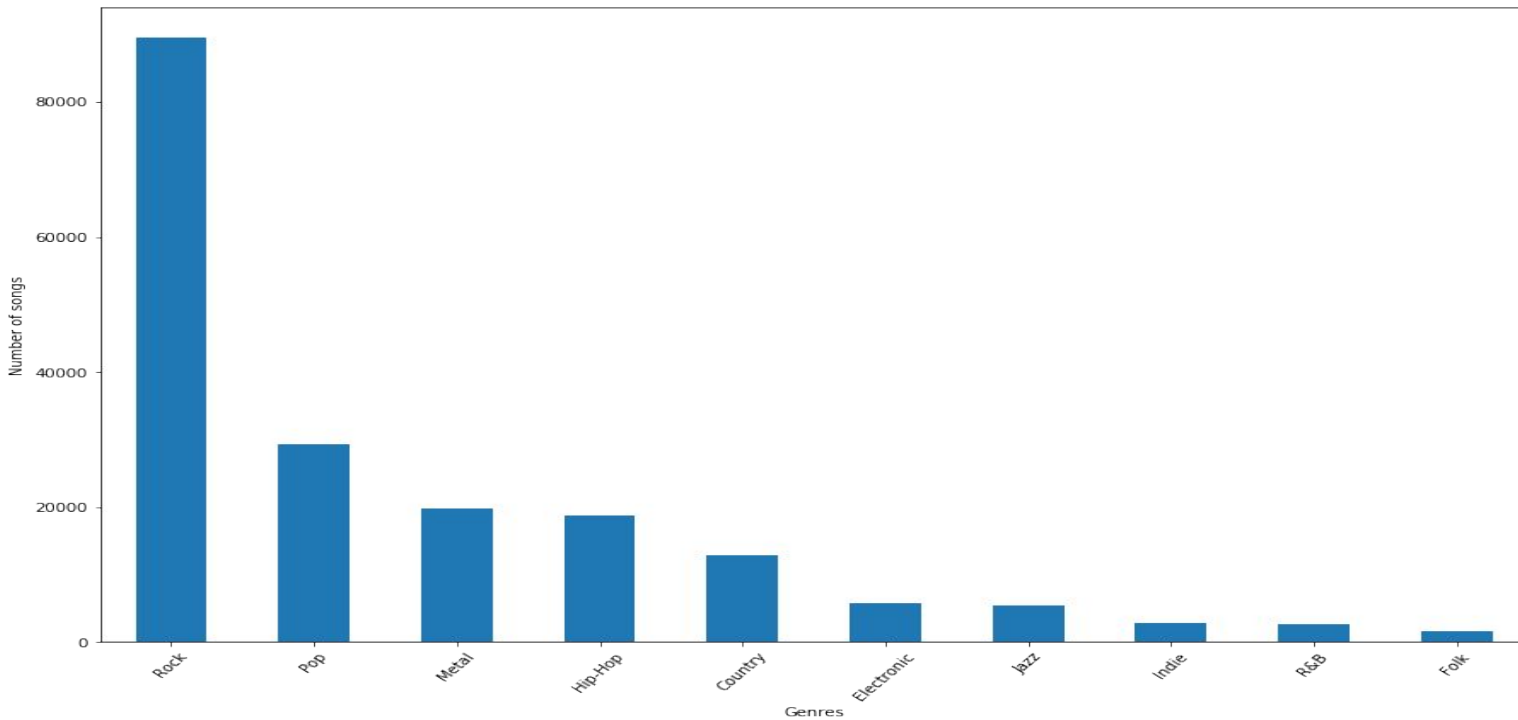
# The Data

Dataset from kaggle 380,000+ lyrics from Metro Lyrics of 10 different Genres with 6 columns: Index, Song title, genre, year, artist, lyric.

Not All the song are English

Unbalancing data between genres

# The Data



# Data Preprocessing

Removing Non-English Songs

Removing songs missing information: “Not Applicable”, “Other” genres, song without lyrics...

Tokenize, lemmatization, standardize format and remove stop words

Balancing the data

# GENRE PREDICTION

Extracting the features: TF-IDF of the lyrics, song titles

Building model using: Multinomial Naive Bayes, Logistic Regression, Neural Network, Grid cross-validation to tune the hyperparameter



# GENRE PREDICTION

	TF-IDF Lyrics	TF-IDF Title 	TF-IDF Lyrics + Title
Multinomial Naïve Bayer	0.4235	0.278	0.3922
Logistic Regression	0.4257	0.2742	0.404
Neural Network	0.414	N/A	N/A

# GENRE PREDICTION

predicted label	Country	240	20	58	5	40	57	4	43	40	58
	Electronic	6	85	16	11	16	4	10	24	10	17
	Folk	17	6	109	1	17	11	2	6	9	16
	Hip-Hop	16	35	14	317	12	22	22	36	34	27
	Indie	13	29	30	4	103	9	12	29	15	34
	Jazz	30	15	23	2	9	165	6	19	35	13
	Metal	5	35	42	9	22	15	289	15	26	34
	Pop	30	99	37	32	60	43	15	137	64	51
	R&B	21	27	10	12	24	50	4	44	122	36
	Rock	22	49	61	7	97	24	36	47	45	114
		Country	Electronic	Folk	Hip-Hop	Indie	Jazz	Metal	Pop	R&B	Rock
		true label									

# REASON GENERATION

Generated reasons to support the prediction

Features like most common repeated words, parts of speech, unique words etc are used

Dynamic Selection of Features

At least one reason to support

More pronouns in Hip-hop

Repeated words in more than 1 genre

# SIMPLE RECOMMENDER SYSTEM

Enormous data has been randomly sampled

Tf-idf vectors extracted including n-grams (1-3)

Cosine similarity scores calculated

# USER INTERFACE

Django Framework

Model views controller

Two pages web app

Entering lyrics

Results display

DEMO

# Challenges

Messy, heavy skew data

Complicated relation between different genre in music (Such as Inde genre can be Indie Rock or Indie Pop)

Training time

# CONCLUSION/FUTURE WORKS

Song Lyrics and Genre are not highly correlated

Same is the case with song title and genre

Predictions of genre

Reasoning for prediction

Recommendation based on the lyrics

Use CNN/ Attention networks as future work