Lyrical Analysis of Rap Misogyny

Michael Cao, Michael Wang, Indy Gu

Background

The influence of American rap can be felt around the world. This **cultural export** raises important concerns about the role of rap & hip hop culture on the spread of misogyny in pop culture. We want to evaluate the truth of some assumptions that are commonly held, using textual data:

- **Stereotypes:** Is rap music really as misogynistic as is commonly believed?

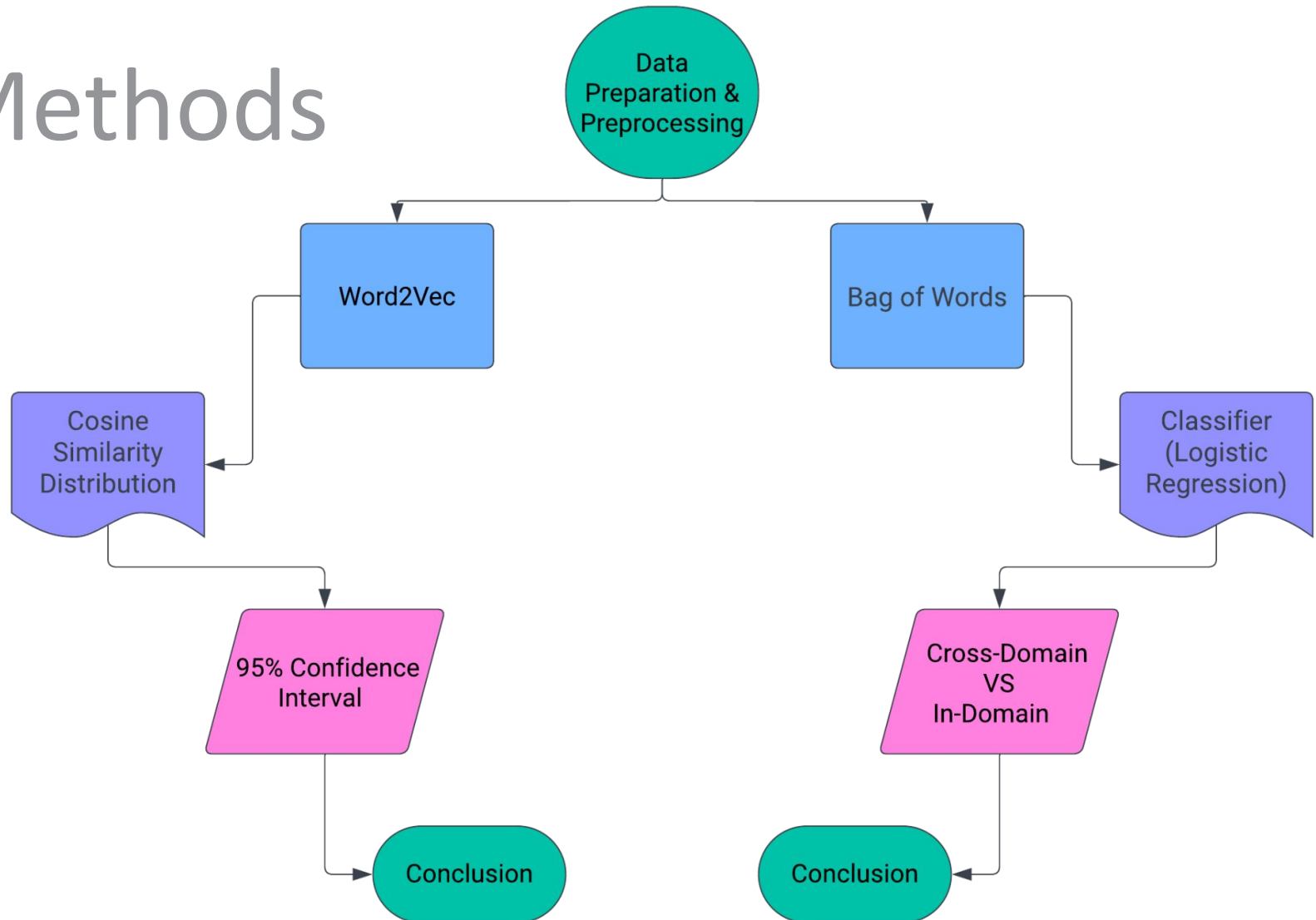- **Diverse Interpretations:** Can the prevalence of misogyny in the music industry be attributed to

Data

1. Lyrics: Database of scraped lyrics Rap-genred songs on Kaggle.

- **Rap Song Dataset:** 4,578 lyrics; Rappers, Verses.

- **General Lyrics Dataset:** 1,840 lyrics; Artists, Genres, Lyrics.

2. Misogynistic Words: Dataset of Reddit posts from traditionally misogynistic communities. Previous researchers have conducted labeling and annotation tasks to determine if the posts are misogynistic or not.

- **Labeled Dataset:** 6,567 Reddit posts ; "Misogynistic" or "Non-Misogynistic".

Methods

Data Preparation & Preprocessing

Word2Vec

Bag of Words

Cosine Similarity Distribution

Classifier (Logistic Regression)

95% Confidence Interval

Cross-Domain VS In-Domain

Conclusion

Conclusion

Diagram of our Methods of Analysis
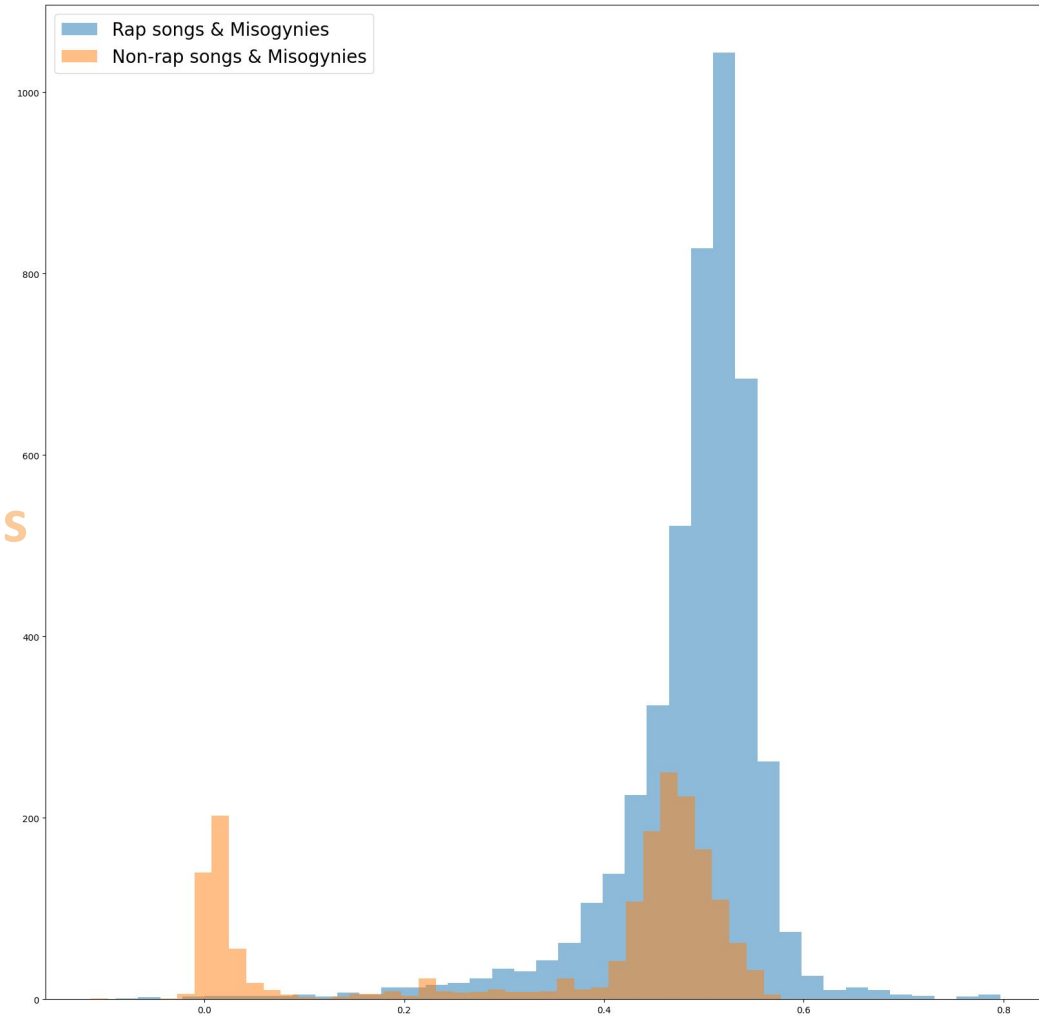
Cosine Similarity Scoring

Confidence Intervals:

Rap Song & Misogynies
- (0.4871, 0.4915)

Non-Rap Songs & Misogynies
- (0.3362, 0.3548)

Domain-Based Predictions

We were able to apply our dataset of misogynistic words to our rap lyrics via **cross-domain** and **in-domain** prediction to obtain a broader scope:

	precision	recall	f1-score	support
Misogynistic	0.79	0.54	0.64	147
Nonmisogynistic	0.94	0.98	0.96	1167
accuracy			0.93	1314
macro avg	0.87	0.76	0.80	1314
weighted avg	0.93	0.93	0.93	1314

Cross-Domain Prediction Results

	precision	recall	f1-score	support
Misogynistic	0.57	0.57	0.57	7
Nonmisogynistic	0.77	0.77	0.77	13
accuracy			0.70	20
macro avg	0.67	0.67	0.67	20
weighted avg	0.70	0.70	0.70	20

In-Domain Prediction Results
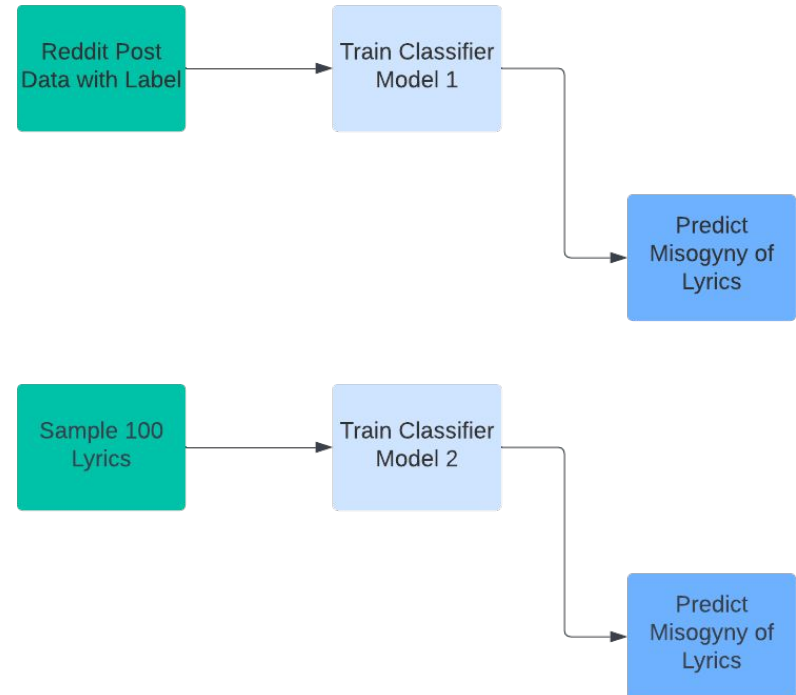
Limitations

Some examples of technical limitations which should be considered for further analysis are as follows:

- **Word2Vec Limitations with Slang:** Word2Vec is unable to fully understand slang and culturally-specific language.

- **Sole Use of Logistic Regression Classifier:** Reliance only on Logistic Regression Classifier for analysis.

- **Subjectivity in Misogynistic Interpretation:** Variability in interpreting and labeling lyrics as misogynistic.

- **Data Constraints:** Limited size of the dataset + lack of comprehensive labeled lyric data.

Complexities to Consider

There are several intriguing social issues and aspects of language that we should consider:

- **Cultural Impact:** Rap music's broad influence sparks important discussions about misogyny in pop culture.

- **Language Nuances:** Rap's rich linguistic structure, utilizing slang, metaphors, and filler words, makes identifying misogyny challenging.

- **Context Sensitivity:** Misogynistic interpretations vary with context.

Limitations
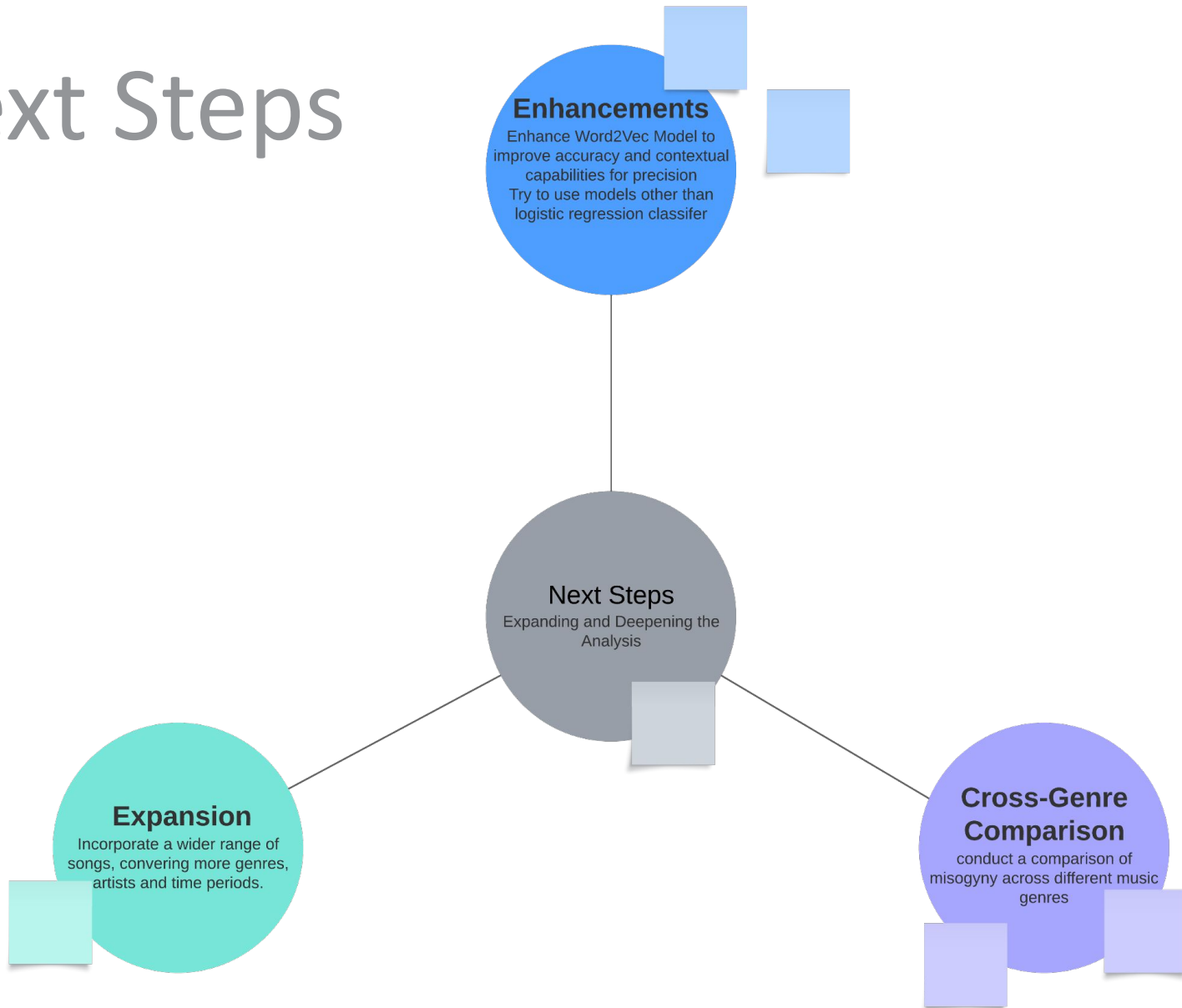
Thank You!
Q&A

Appendix

Rap song lyrics data:
https://github.com/fpaupier/RapLyrics-Scraper/tree/master/lyrics_US

General lyrics data:
https://www.kaggle.com/datasets/neisse/scrapped-lyrics-from-6-genres

Labeling data:

Coates, Joshua, and Danushka Bollegala. "Frustratingly easy meta-embedding –

computing meta-embeddings by averaging source word embeddings." *Proceedings*

of the 2018 Conference of the North American Chapter of the Association for

Computational Linguistics: Human Language Technologies, Volume 2 (Short

Papers), 2018, https://doi.org/10.18653/v1/n18-2031.