PRINCETON UNIVERSITY

Kaur or Kendrick: An Exploratory Analysis of the Overlap Between Rap and Modern Poetry

Kyle Barnes, Roopa Ramanujam, Saahil Katyal

INTRODUCTION (A HAIKU)

rap or poetry?
where does the distinction lie?
use machine learning

- Ambiguity surrounding whether rap is poetry
- Kendrick Lamar: 2018 Pulitzer Prize
- Build off work of Rhody (2012)

DATA

Summary

Poetry Data Set:

- poetryfoundation.org
- “Contemporary”
- 15,652 poems

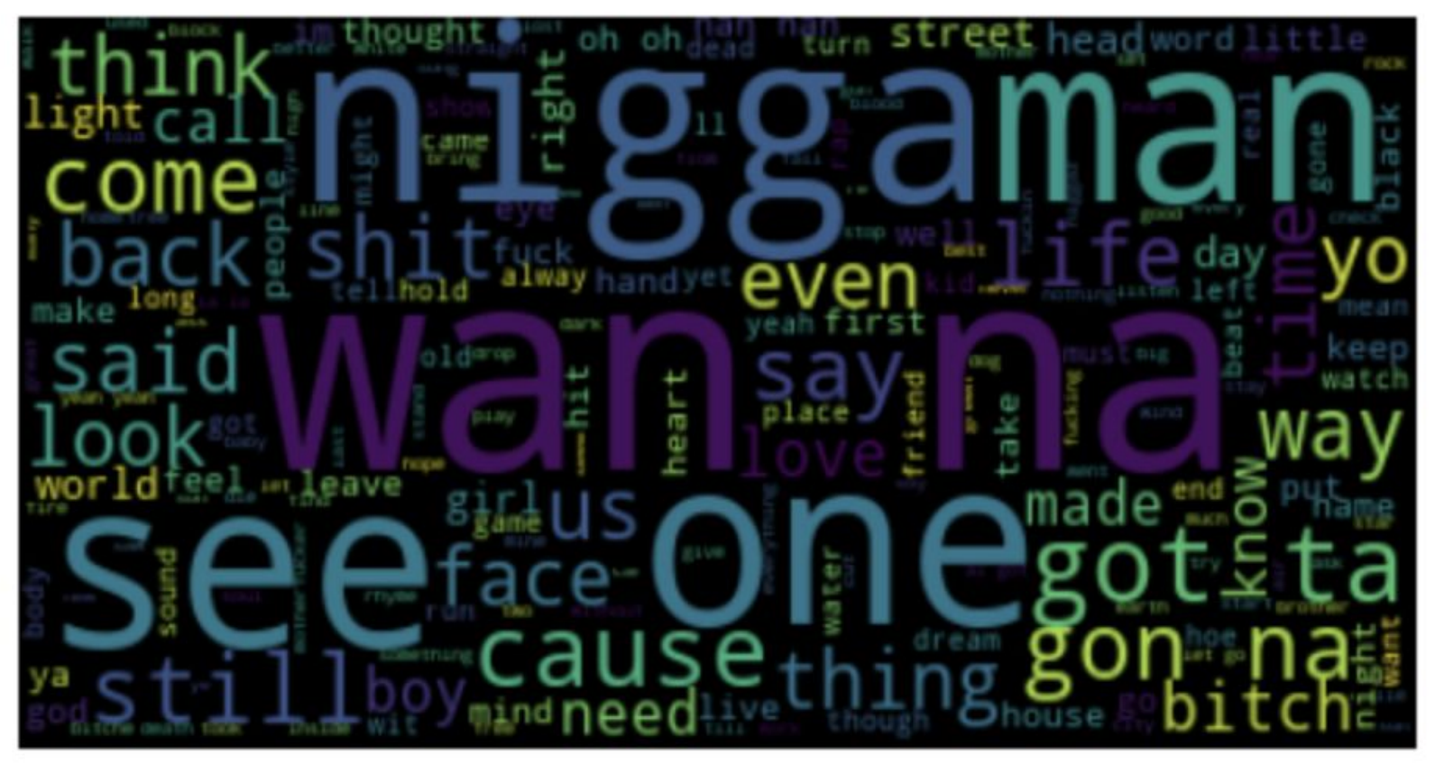
Rap Data Set:

- Metrolyrics - “hiphop”
- 33,965 raps

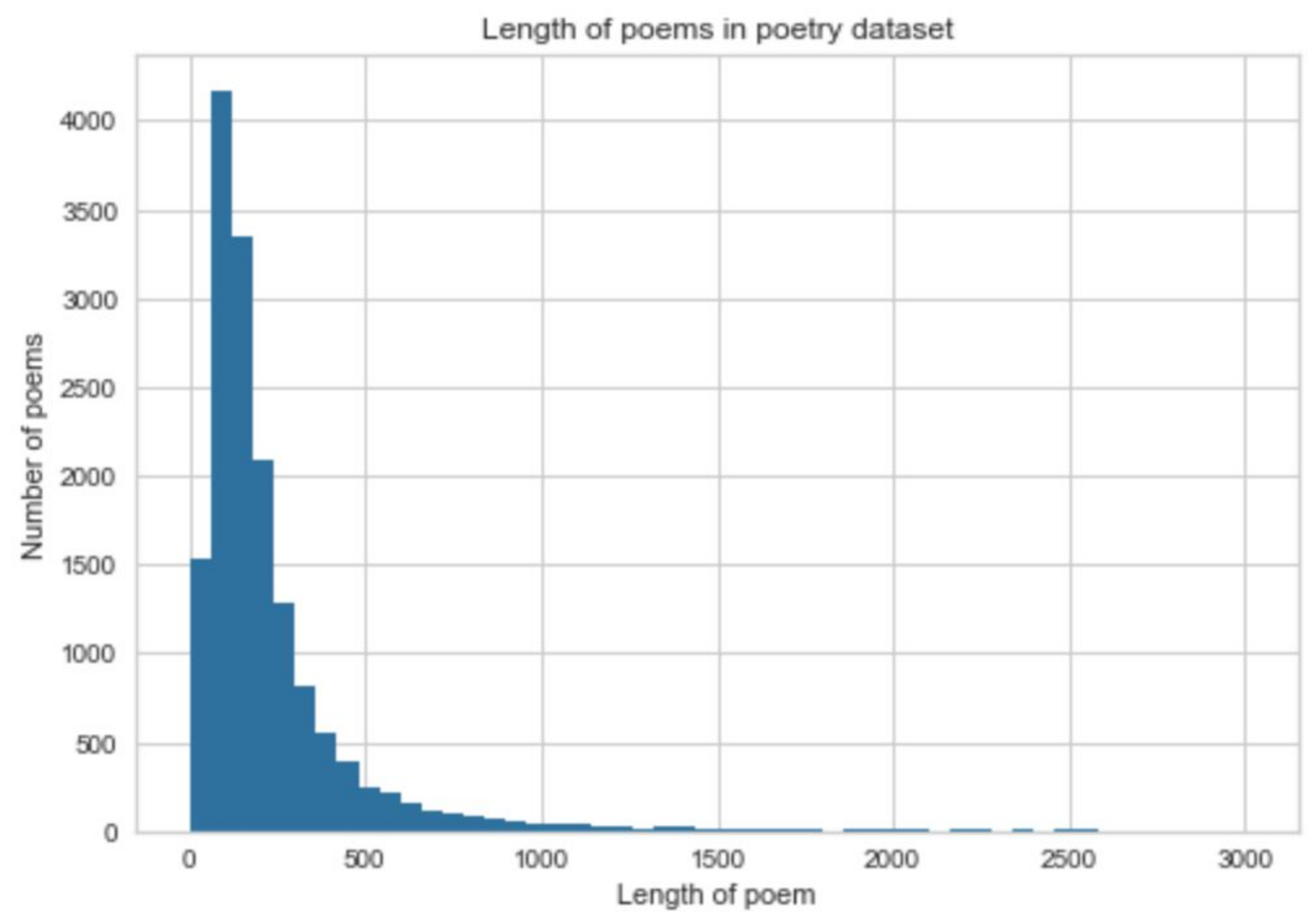
Exploratory Analysis

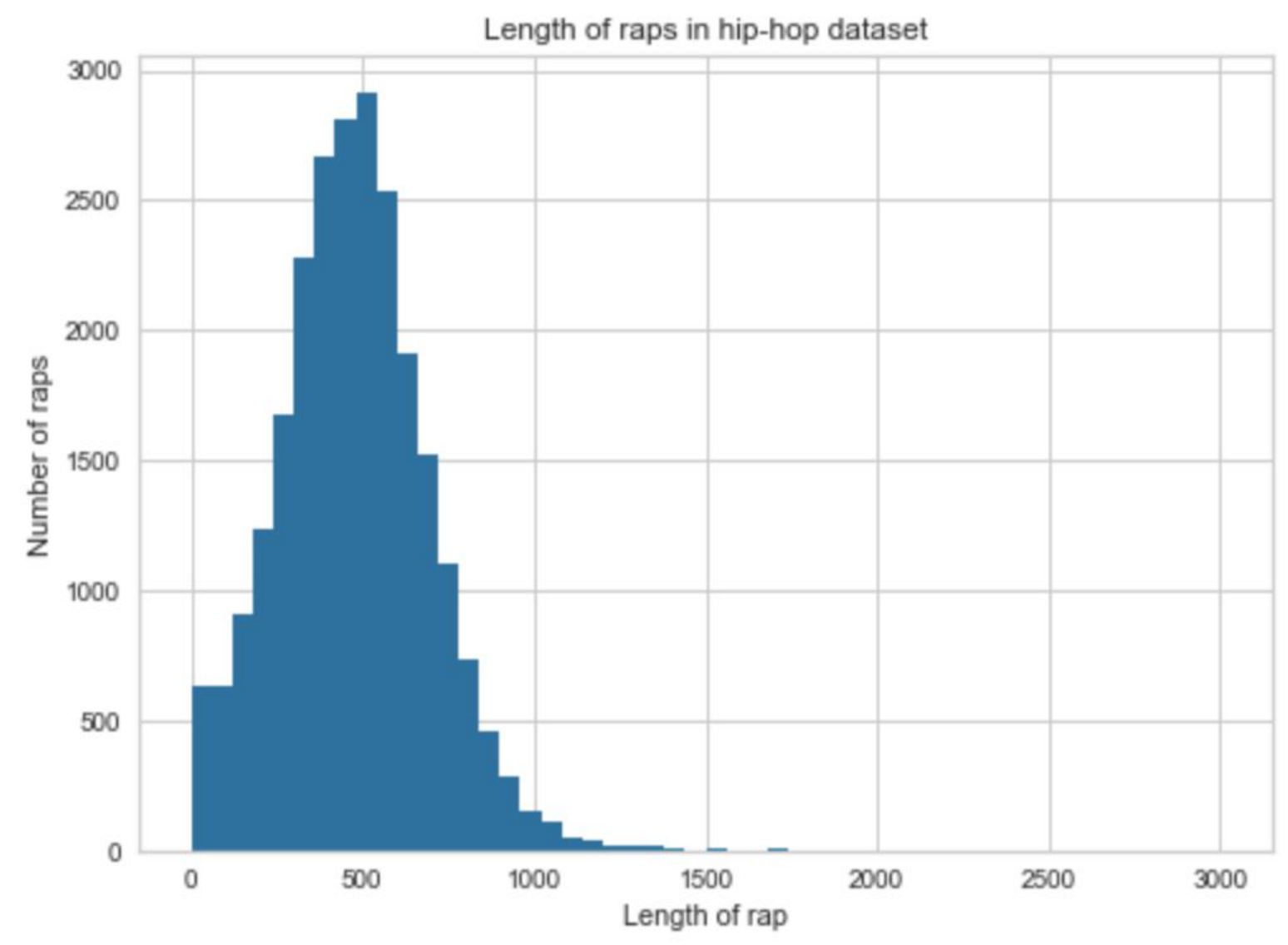
see
love
one
light
time
day
eye
hand
still
life
say
come
heart
world
back
us
water

Poetry Data Set World Cloud



Rap Data Set World Cloud



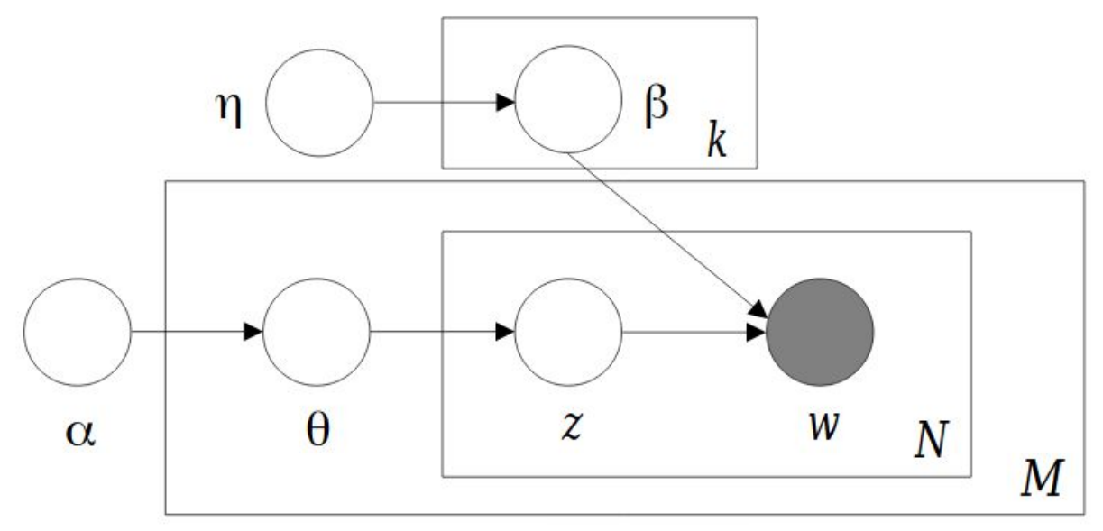


Data Cleaning and Representation

- {Author, Title, Content, Label}
- Dropped NAs and removed html tags and other markers such as “Verse 1”
- Split data set into train and test using 80/20 split
- Tokenized content using bag-of-words and tf-idf representations

METHODS

Latent Dirichlet Allocation Graphical Representation



LDA Model Selection

Number of Topics	Log-Likelihood	Perplexity
2	-37.687	0.0254
3	-38.359	0.0286
4	-38.542	0.0295

Comparing LDA with 2 Components on Tf-Idf and Bag-of-Words

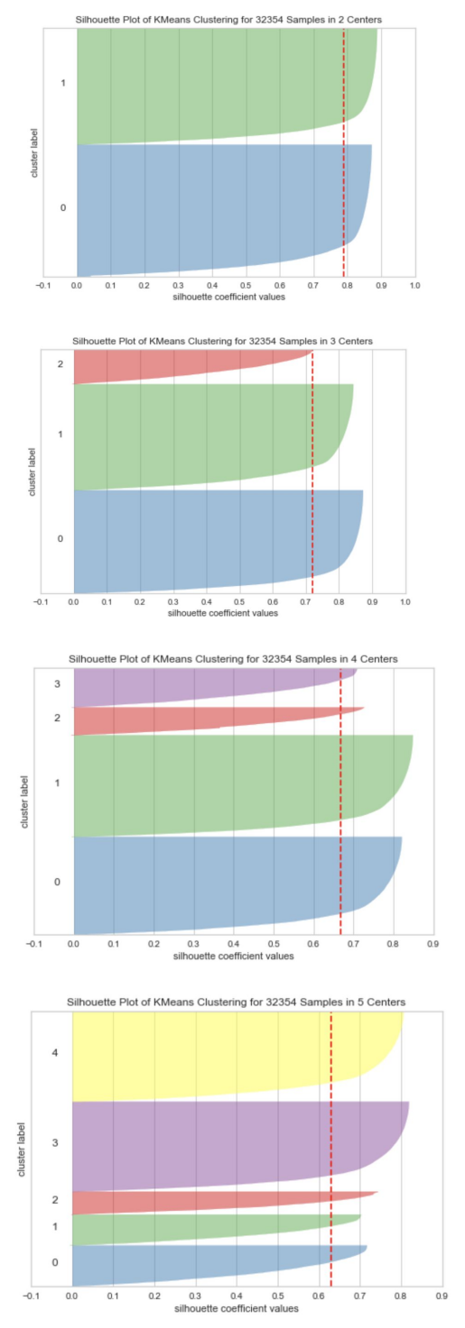
	Tf-Idf	Bag-of-Words
Log-Likelihood	-37.687	-690.59
Perplexity	0.0253	0.0158

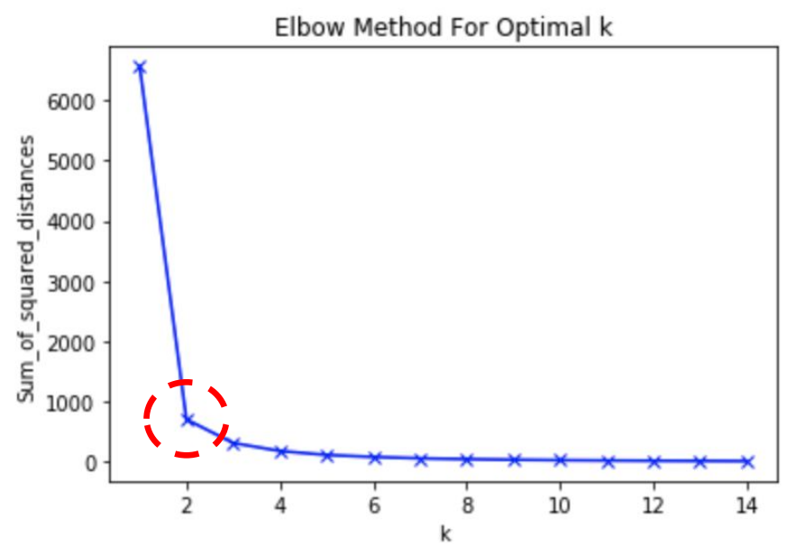
Supervised Learning

- Logistic regression, random forest, naive bayes, support vector classifier, linear support vector classifier

K-Means Clustering on Topic Proportions

- Selected k = 2 clusters based on silhouette and elbow plots





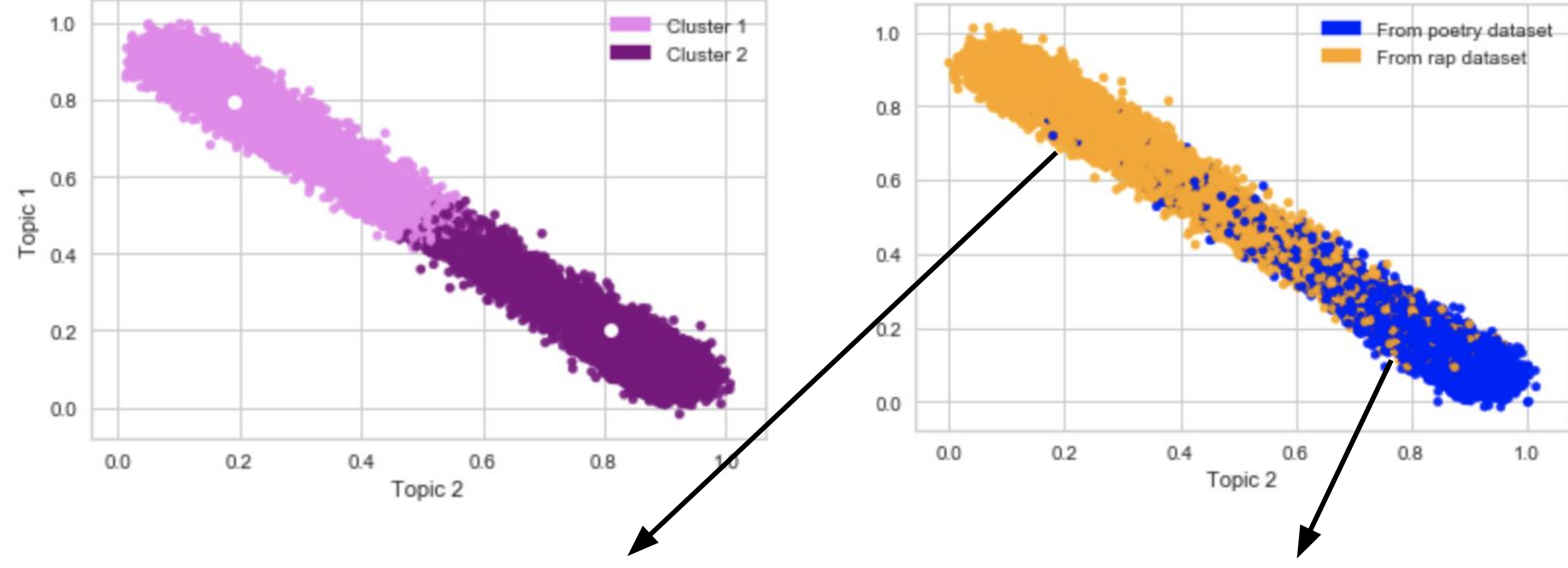
RESULTS

LDA Topics

Topic Number	Label	Tokens
1	rap	like, got, don, know, niggas, just, ain, cause, shit, niggas, ya, ll, fuck, yeah, want, let, baby, yo, make, em
2	poetry	like, love, light, night, day, eyes, world, life, away, long, old, heart, time, white, water, sun, said, amp, god, dark

	Log-Likelihood	Perplexity
Train	-37.687	0.0254
Test	-38.073	0.107

K-Means Clustering



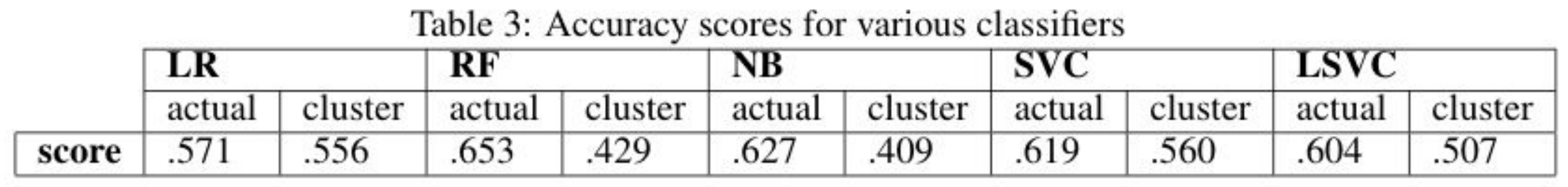
Vignette: Dorothy Parker - Resumé

Razors pain you;
Rivers are damp;
Acids stain you;
And drugs cause cramp.
Guns aren't lawful;
Nooses give;
Gas smells awful;
You might as well live.

Vignette: Drake - Where Were You

Why am I in bed alone?
How come when I drive by,
it looks like you are never
home?
Do you even live there?
Or did you take a U-Haul?
Too far, move off and not
even tell me that you're
gone?

Table 3: Accuracy scores for various classifiers



	LR		RF		NB		SVC		LSVC	
	actual	cluster	actual	cluster	actual	cluster	actual	cluster	actual	cluster
score	.571	.556	.653	.429	.627	.409	.619	.560	.604	.507

ACKNOWLEDGEMENTS

COS 424 course staff

REFERENCES

1. Rhody, L. Topic Modeling and Figurative Language. In: Journal of Digital Humanities; 2012.
2. ultra-jack. Poems from PoetryFoundation.org: Modern and Renaissance Poetry for Classification Exercises. Kaggle; 2017. https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics.
3. Mishra, G. 380,000 Lyrics from MetroLyrics. Kaggle; 2017. https://www.kaggle.com/ultrajack/modern-renaissance-poetry.