



PREDICTING POETIC MOVEMENTS

BY JOSHUA SZYMANOWSKI

WHY?





A LOOK AT THE DATA

SOURCE

poetryfoundation.org

TOTAL POEMS

5,168 → 4,307

TARGET

movement

YEARS

1780 - present

WORDS

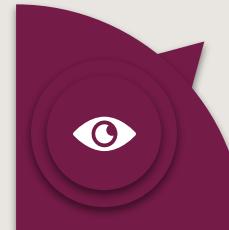
1,165,726 – 119,285 unique

TARGET

Pre-1900
(24.1%)



Modern
(28.9%)



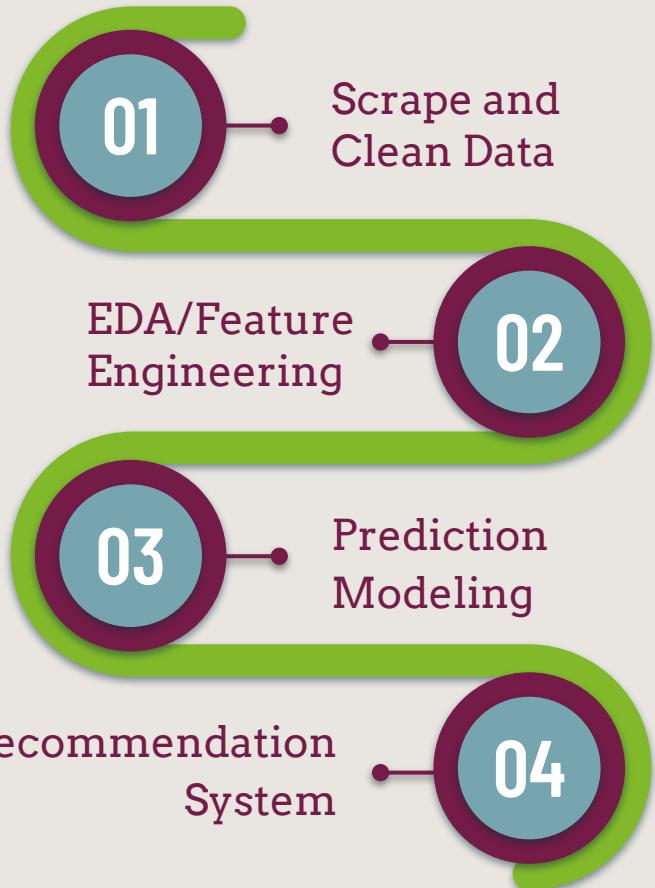
Metropolitan
(24.9%)



Avant-Garde
(22.1%)



ROADMAP



MODEL PREVIEW



ACCURACY

45.9%

F1 SCORE

43.0%



ACCURACY

59.4%

F1 SCORE

59.5%



ACCURACY

65.6%

F1 SCORE

65.5%



ACCURACY

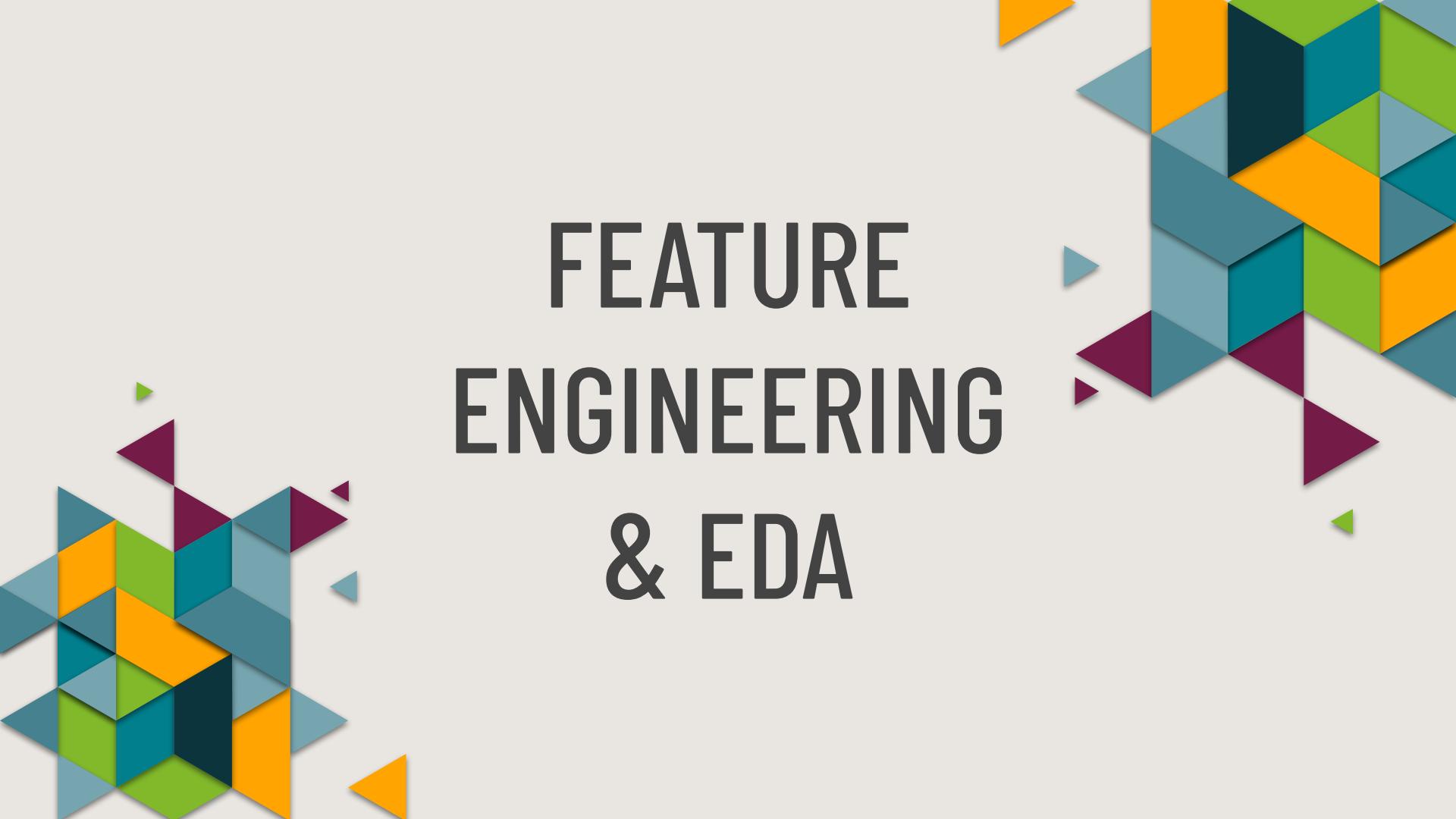
64.4%

F1 SCORE

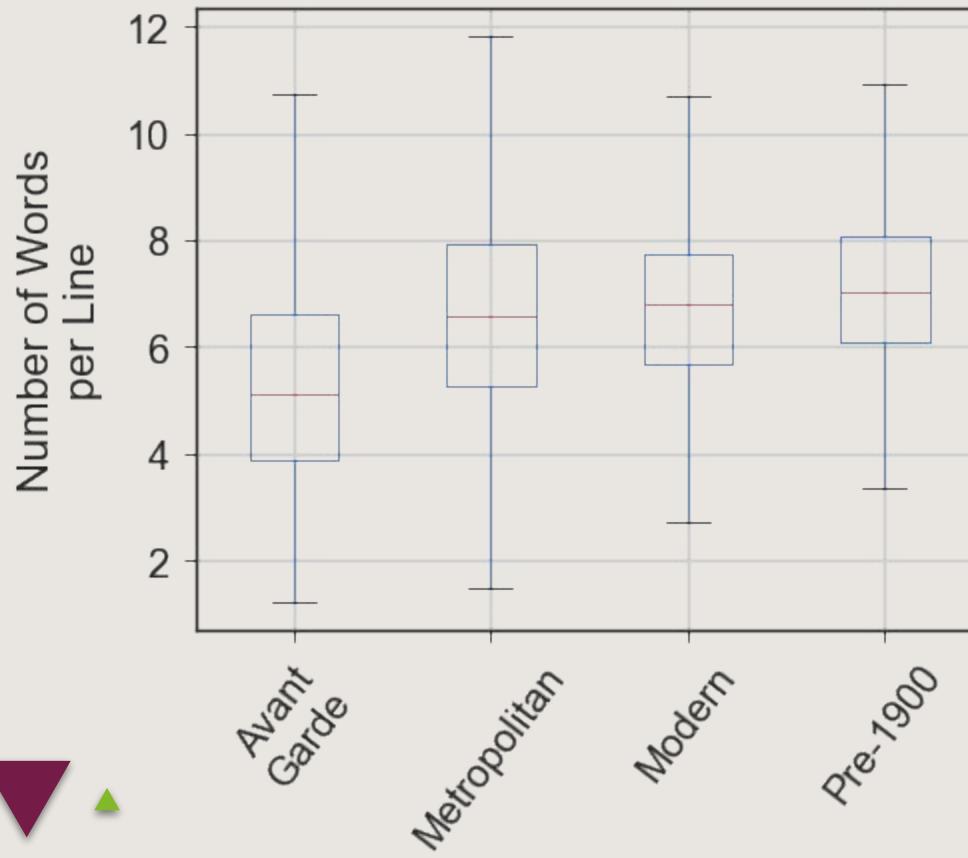
63.9%



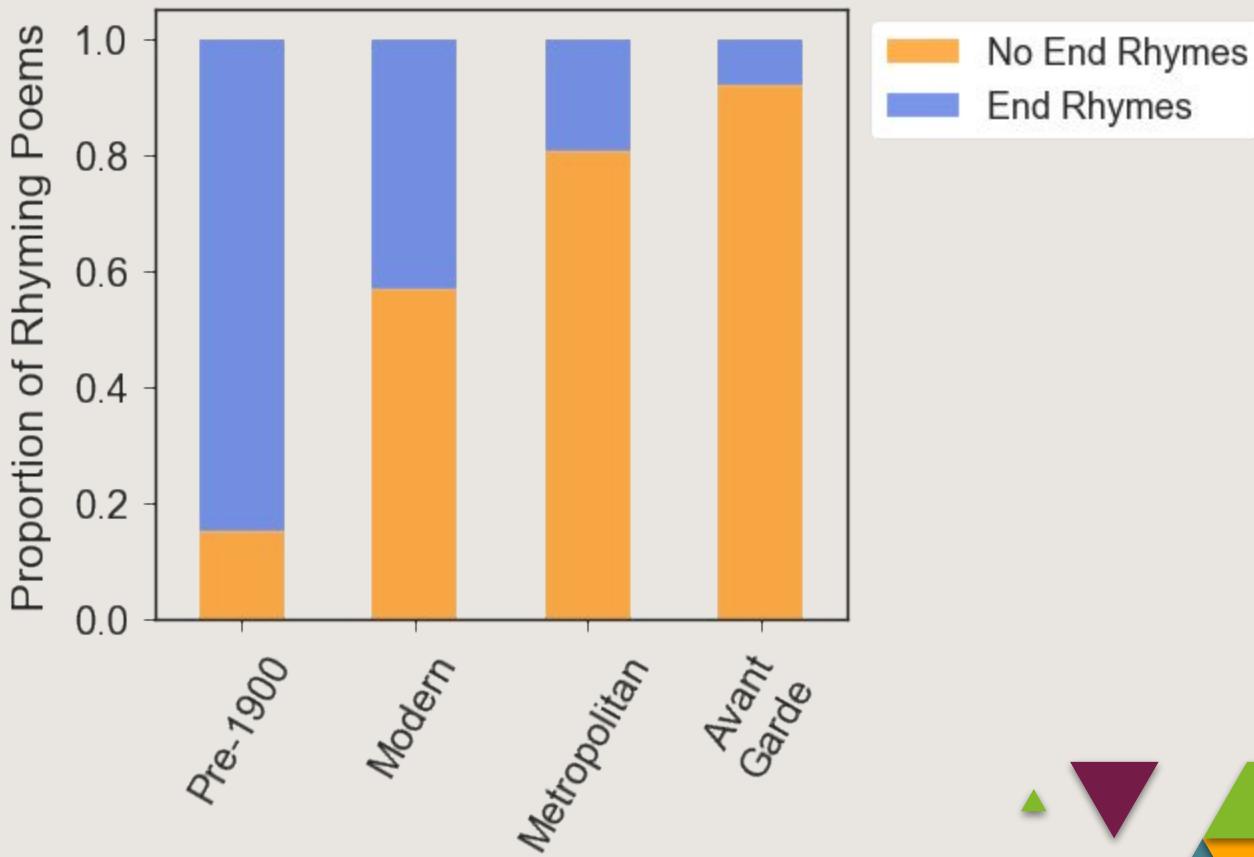
FEATURE ENGINEERING & EDA



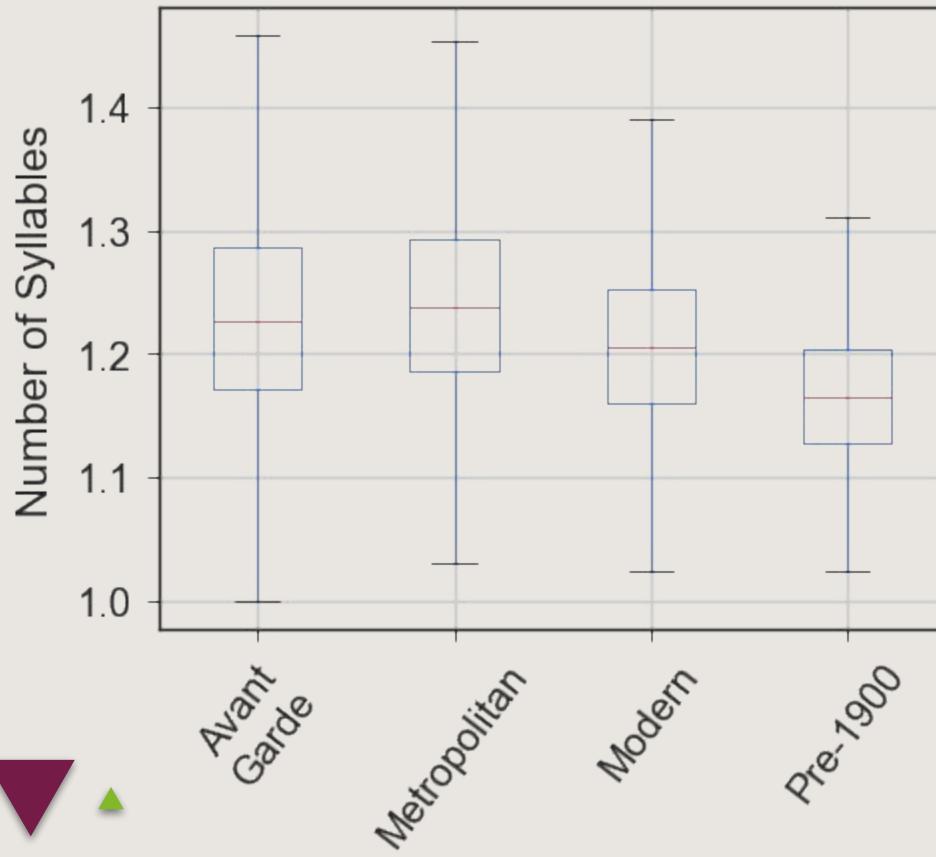
Average Line Length, Distribution by Movement



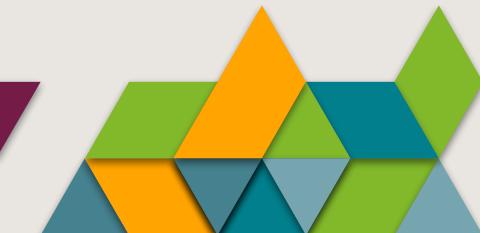
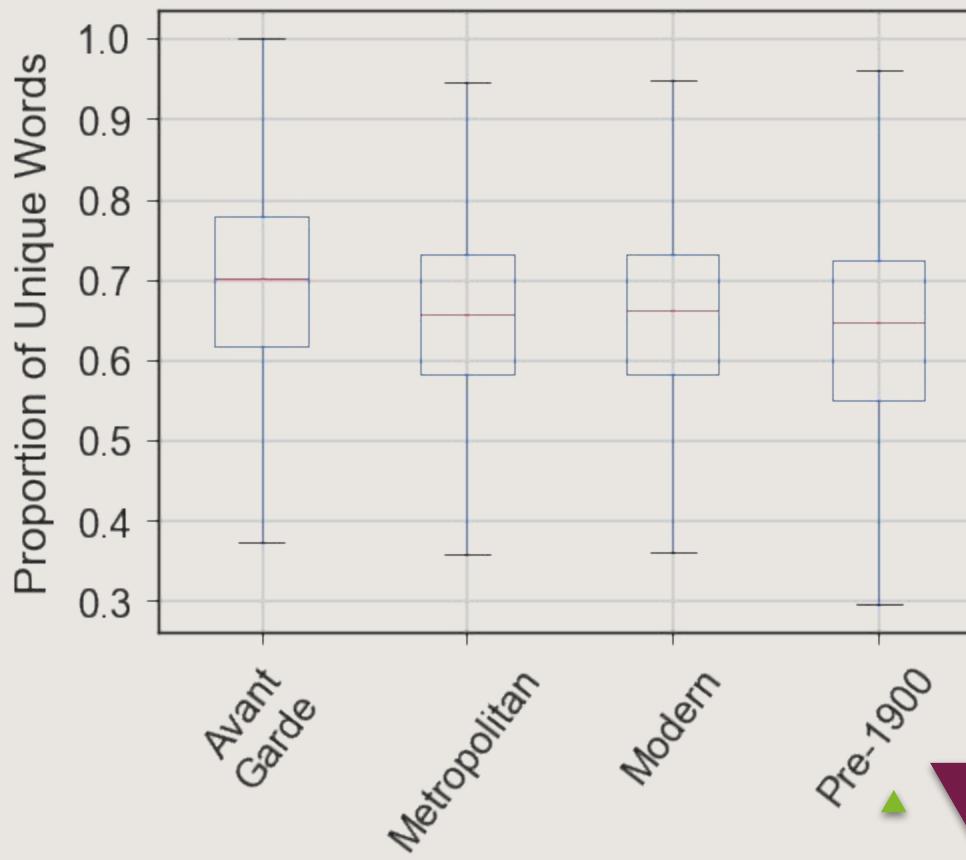
Usage of End Rhymes Across Poetic Movements



Average Number of Syllables per Word, Distribution by Movement

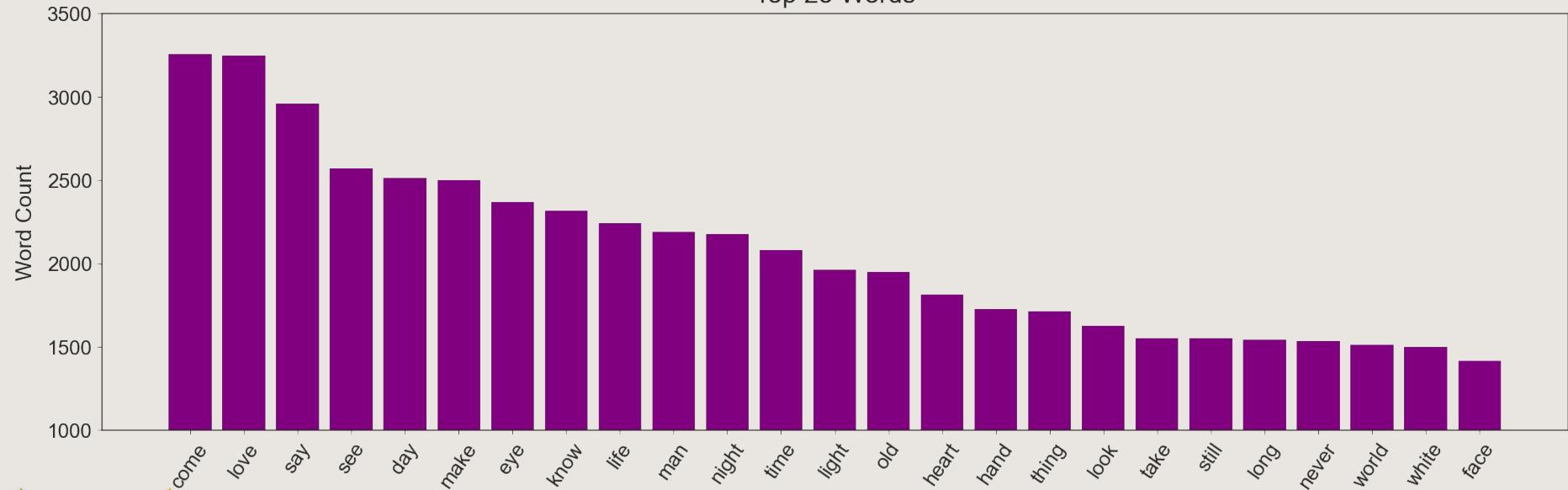


Lexical Richness, Distribution by Movement



Visual | Temporal | Conceptual

Top 25 Words



Metropolitan

Modern

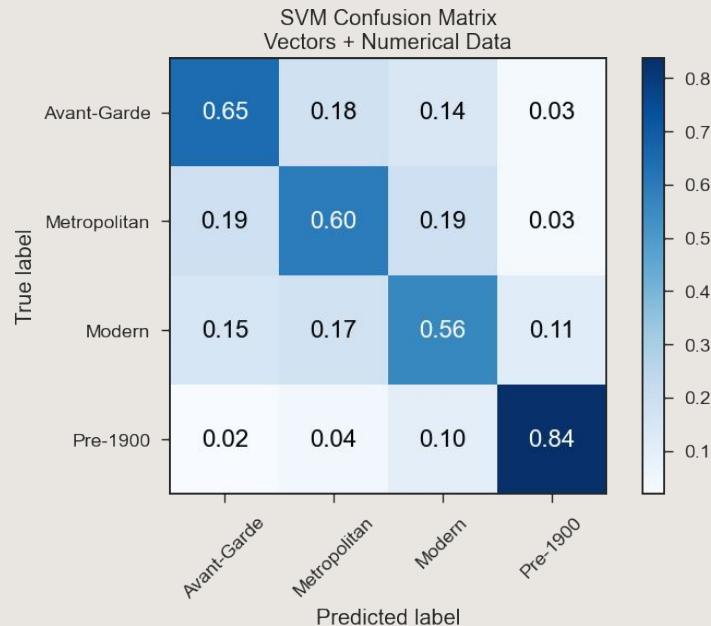
Avant-Garde

Pre-1900

MODELS

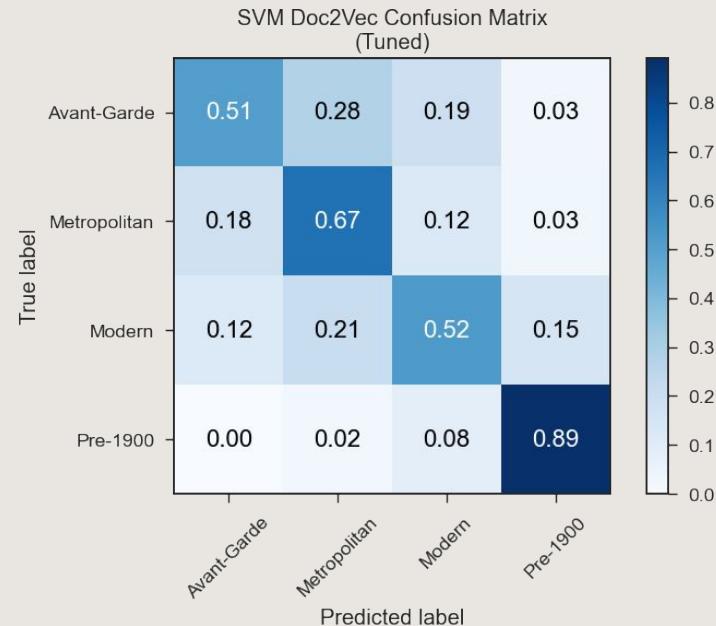


SVM w/ TF-IDF



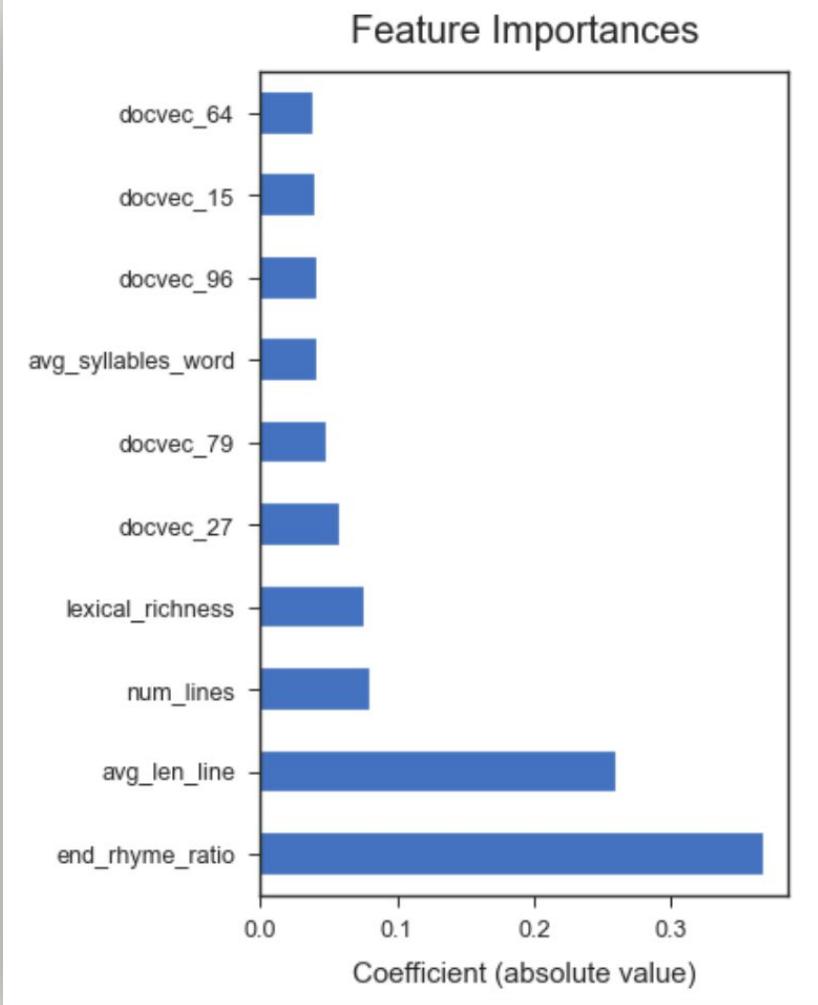
Training F1 99.9%
Testing F1 65.5%

SVM w/ Doc2Vec



Training F1 67.0%
Testing F1 63.9%

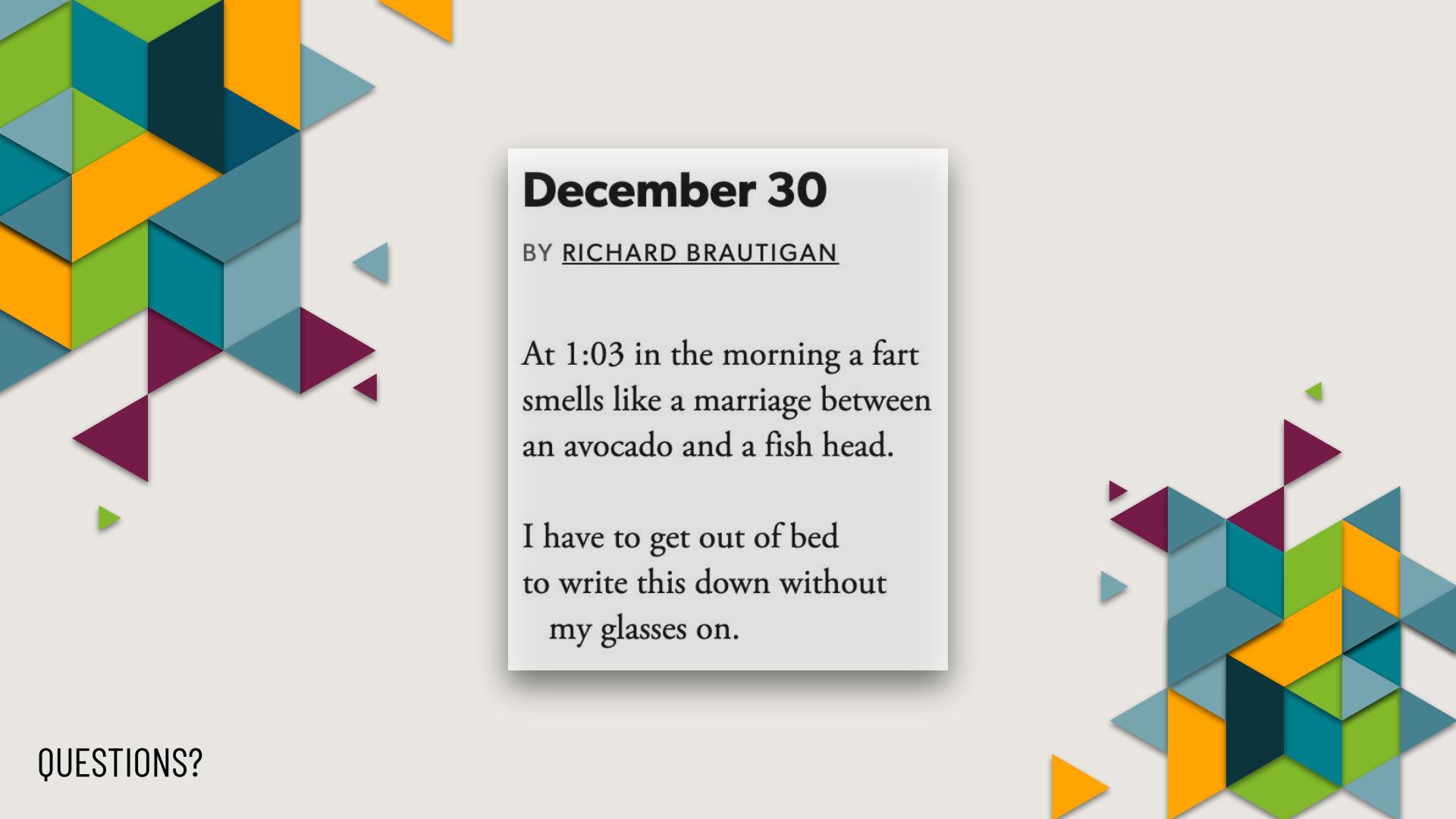
TOP FEATURES





TAKEAWAYS

- Pre-1900 poetry is easily recognized
 - End rhymes and length of line
 - Avant-garde is its polar opposite
 - Poetry is rarely neutral (and generally positive)
 - Form/structure important for most prediction models
 - SVM Doc2Vec provides best fit
- 



December 30

BY RICHARD BRAUTIGAN

At 1:03 in the morning a fart
smells like a marriage between
an avocado and a fish head.

I have to get out of bed
to write this down without
my glasses on.

QUESTIONS?

Joshua Szymanowski

joshua.szymanowski@gmail.com

CONTACT INFO

github

LinkedIn