



PREDICTING POETIC MOVEMENTS

BY JOSHUA SZYMANOWSKI

WHY?





A LOOK AT THE DATA

SOURCE

poetryfoundation.org

TOTAL POEMS

5,176 → 4,376

TARGET

genre

YEARS

1780 - present

WORDS

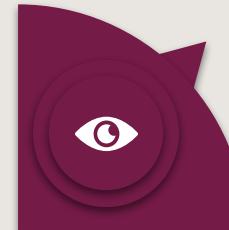
1,034,736 – 110,723 unique

TARGET

Pre-1900
(23.8%)



Modern
(29.2%)



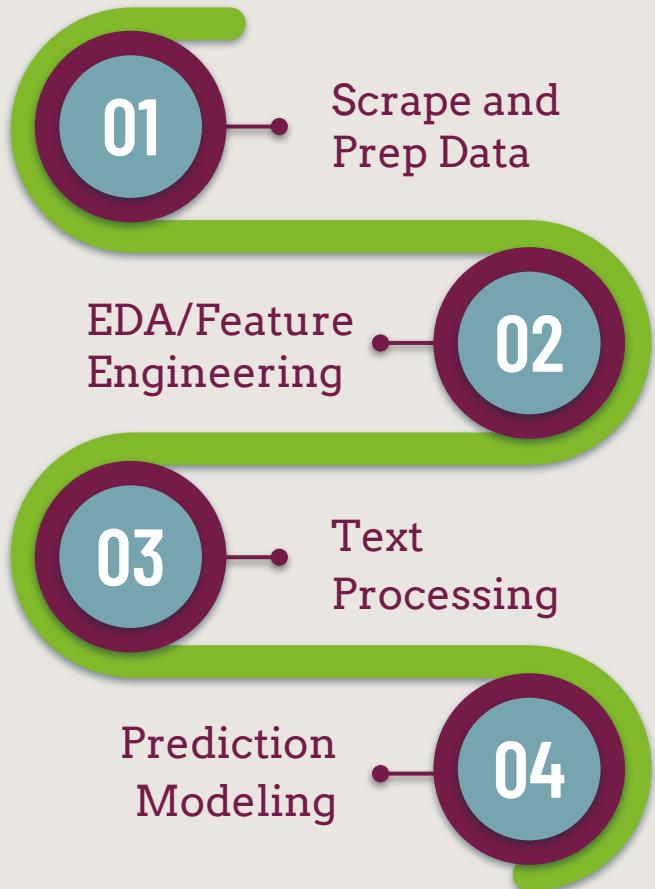
Metropolitan
(25.0%)



Avant-Garde
(22.0%)



ROADMAP



MODEL PREVIEW



ACCURACY

46.7%

F1 SCORE

42.6%



ACCURACY

47.2%

F1 SCORE

47.0%



ACCURACY

62.6%

F1 SCORE

62.1%



ACCURACY

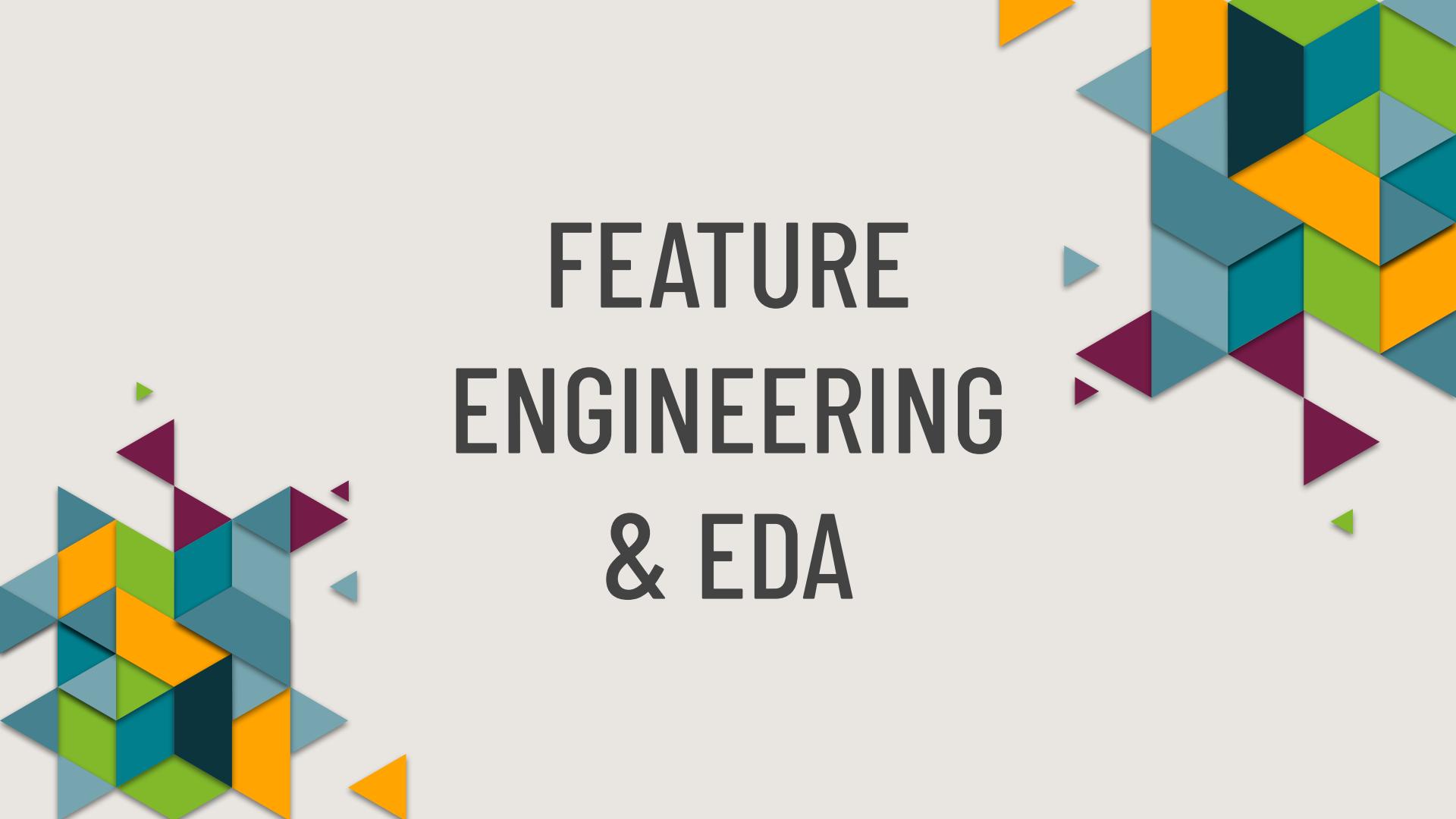
67.8%

F1 SCORE

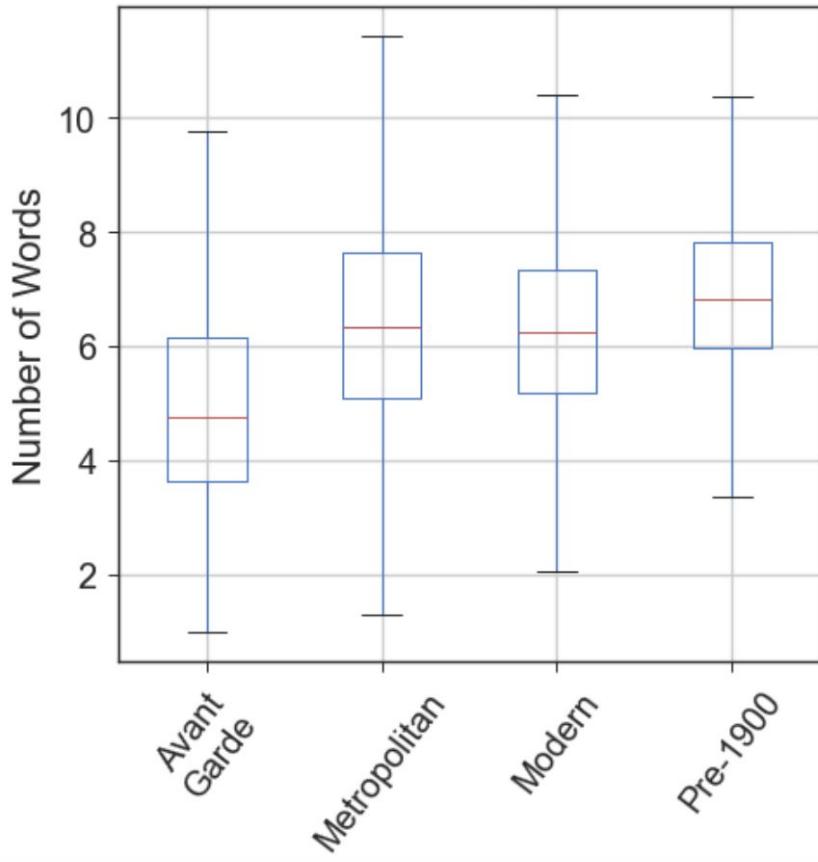
67.8%



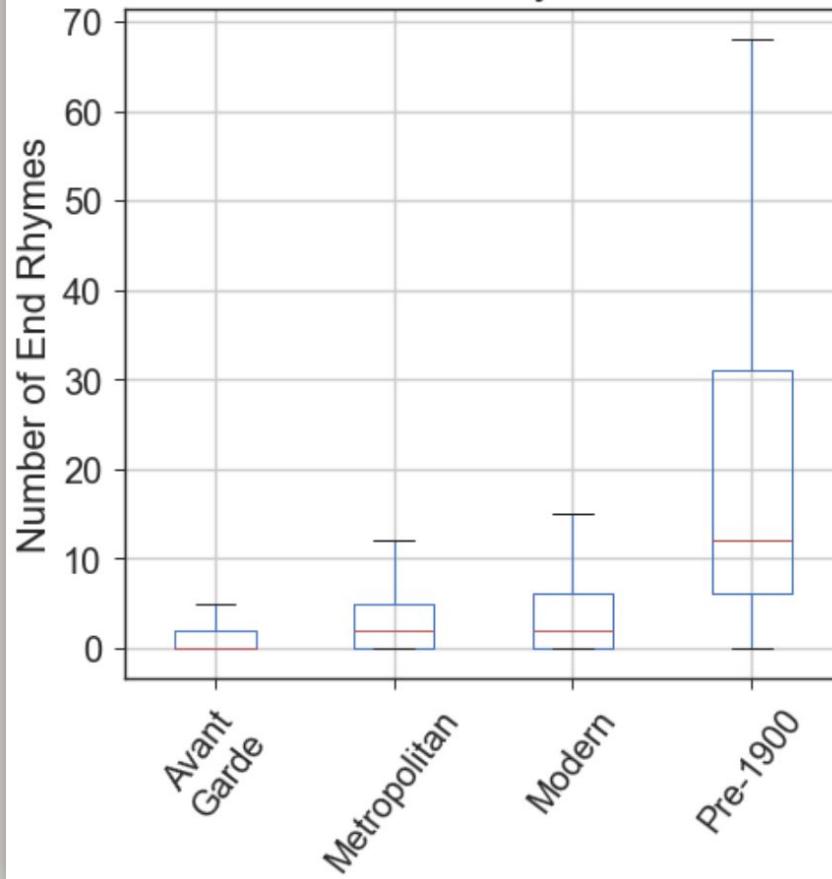
FEATURE ENGINEERING & EDA



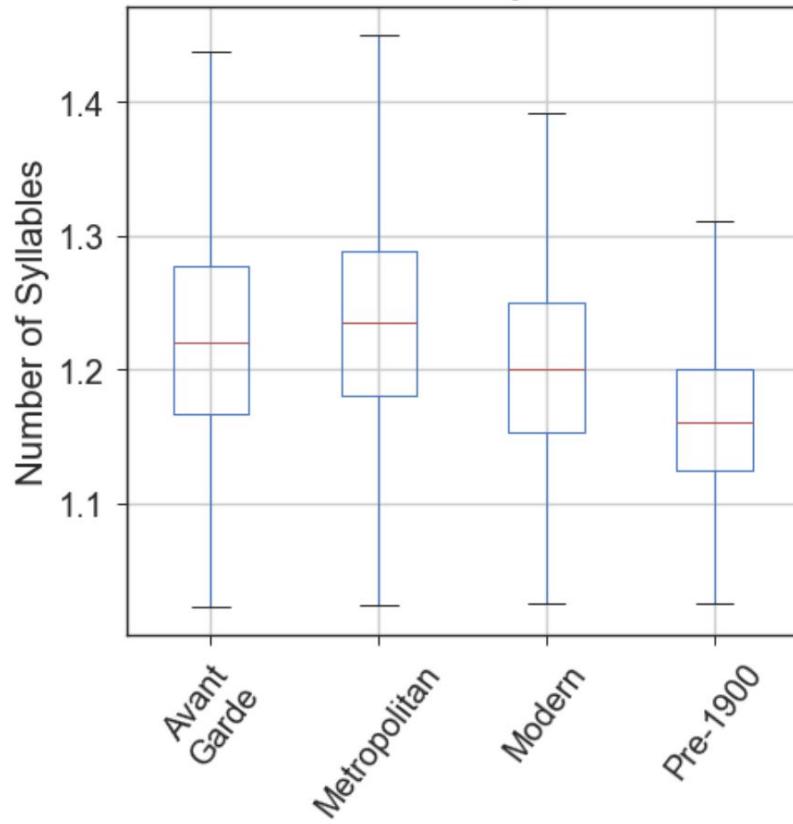
Average Line Length per Poem, Distribution by Genre



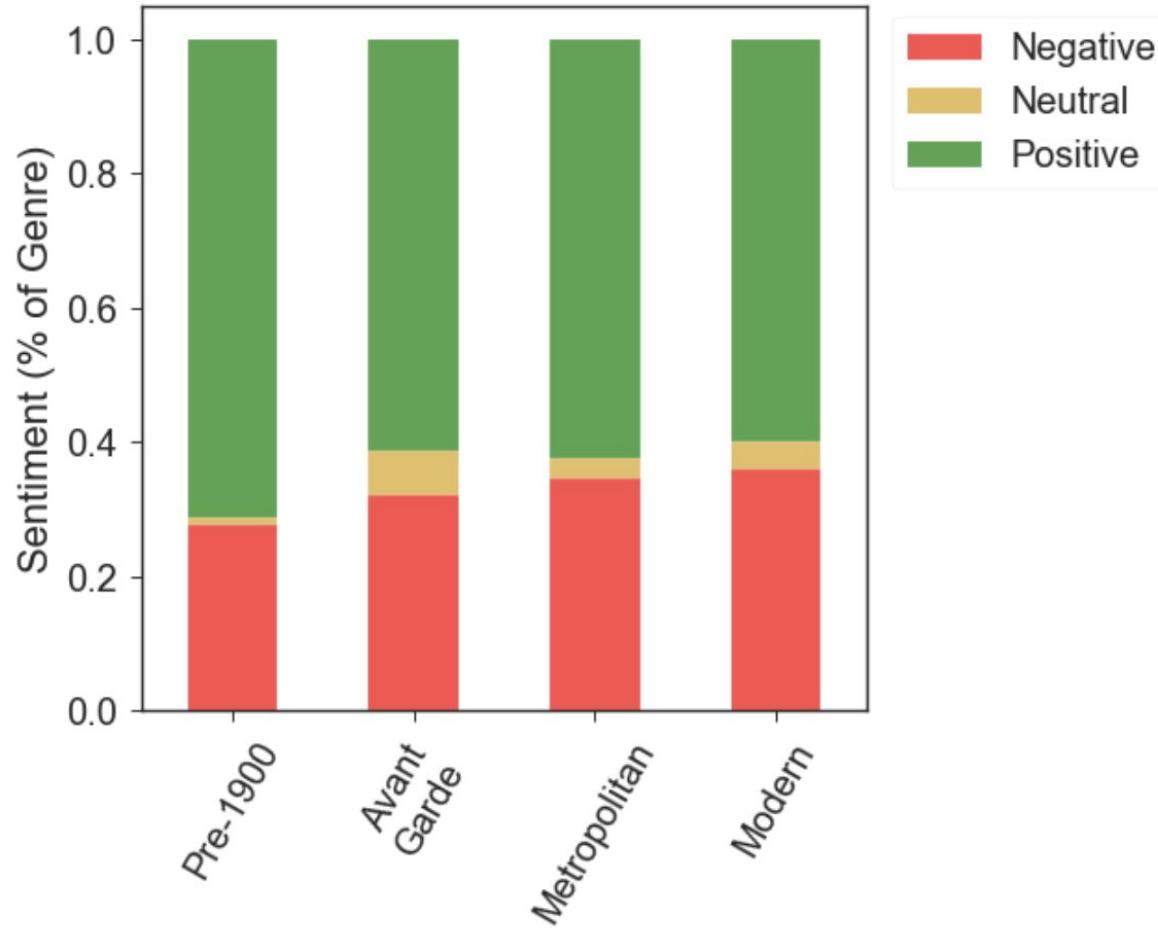
Number of End Rhymes, Distribution by Genre



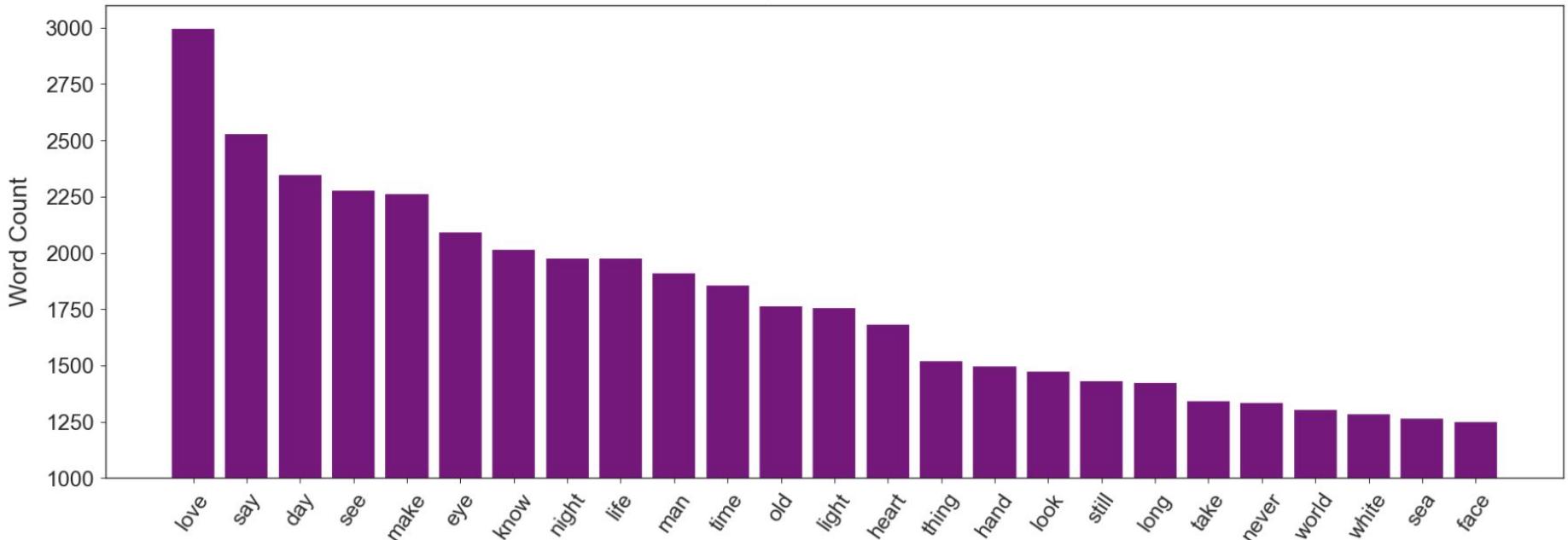
Average Number of Syllables Per Word, Distribution by Genre



The Polarity of Poetic Genres



Top 25 Words



Metropolitan

Modern

people light heart long grow mountain hold silver finger half verse
black mind bird blood flower alone god young color yellow gold last
hour shadow young color stone high set rain another hair
away show left lie grass city live die
moon without keep well thought star stone woman saw
music first round must foot soul till born green cold room street
must first round must foot soul till born green cold room street
work find spring poem year red men still dan song world
find spring poem year red men still dan song world
fall water two thing name strong end smile hill part
water two thing name strong end smile hill part
take move much summer air back wall place fire
move much summer air back wall place fire
night far tree dust something new heard turn land wing
far tree dust something new heard turn land wing
day window whose house sleep wait arm open sky word
window whose house sleep wait arm open sky word
old stand over cloud silence golden home always night
stand over cloud silence golden home always night
look door head earth sun change call hear never little
door head earth sun change call hear never little
death sound heaven beauty great dream leaf

Avant-Garde

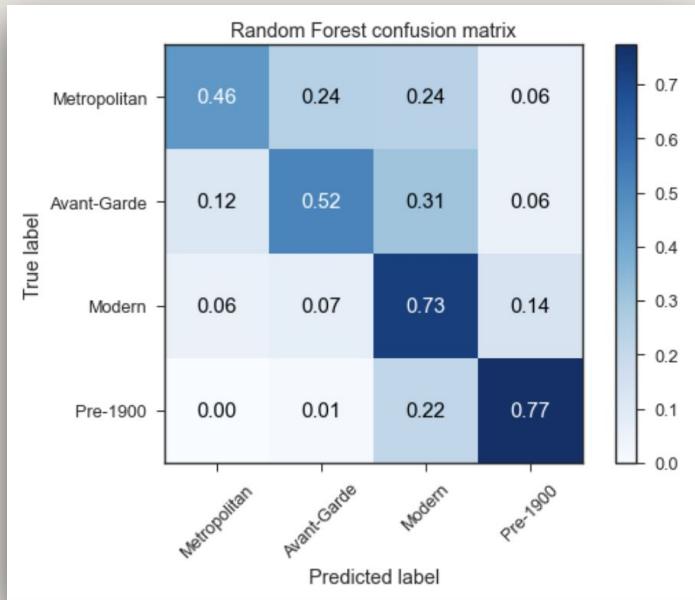
Pre-1900

golden
man part
sweet leaf-wood clear
old poor cry voice star inside
thought ever air must forth
rise wing stood fill
mountain smile Whose great gold
day cold earth spring foot
eye red song white
far bed tell left mind sight
thing night know stream blood
mine till though water power
hand clear leave lie good
sea pale fear move
never set land door
find hour dear
wind change
world field without
look art
take nature blue lay
keep away
hear bright green king
t little fire well soft
see round bird rest
make head sun within
feel tear live word
soul hill full watch ion
time friend half young
say pa-brest since
name wild grow name
saw black snow
light cloud night
fairy kiss child first

MODELS

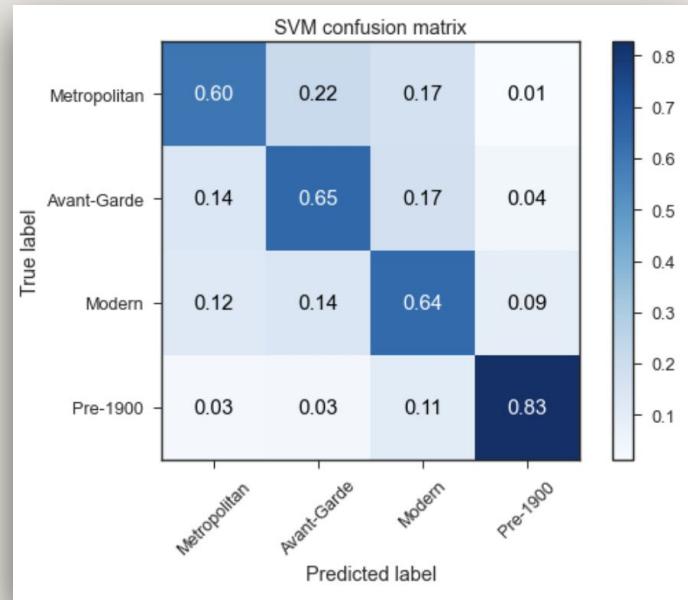


RANDOM FOREST MODEL



ACCURACY 62.6%
F1 SCORE 62.1%

SVM MODEL



ACCURACY 67.8%
F1 SCORE 67.8%

RANDOM FOREST

SVM

TOP FEATURES

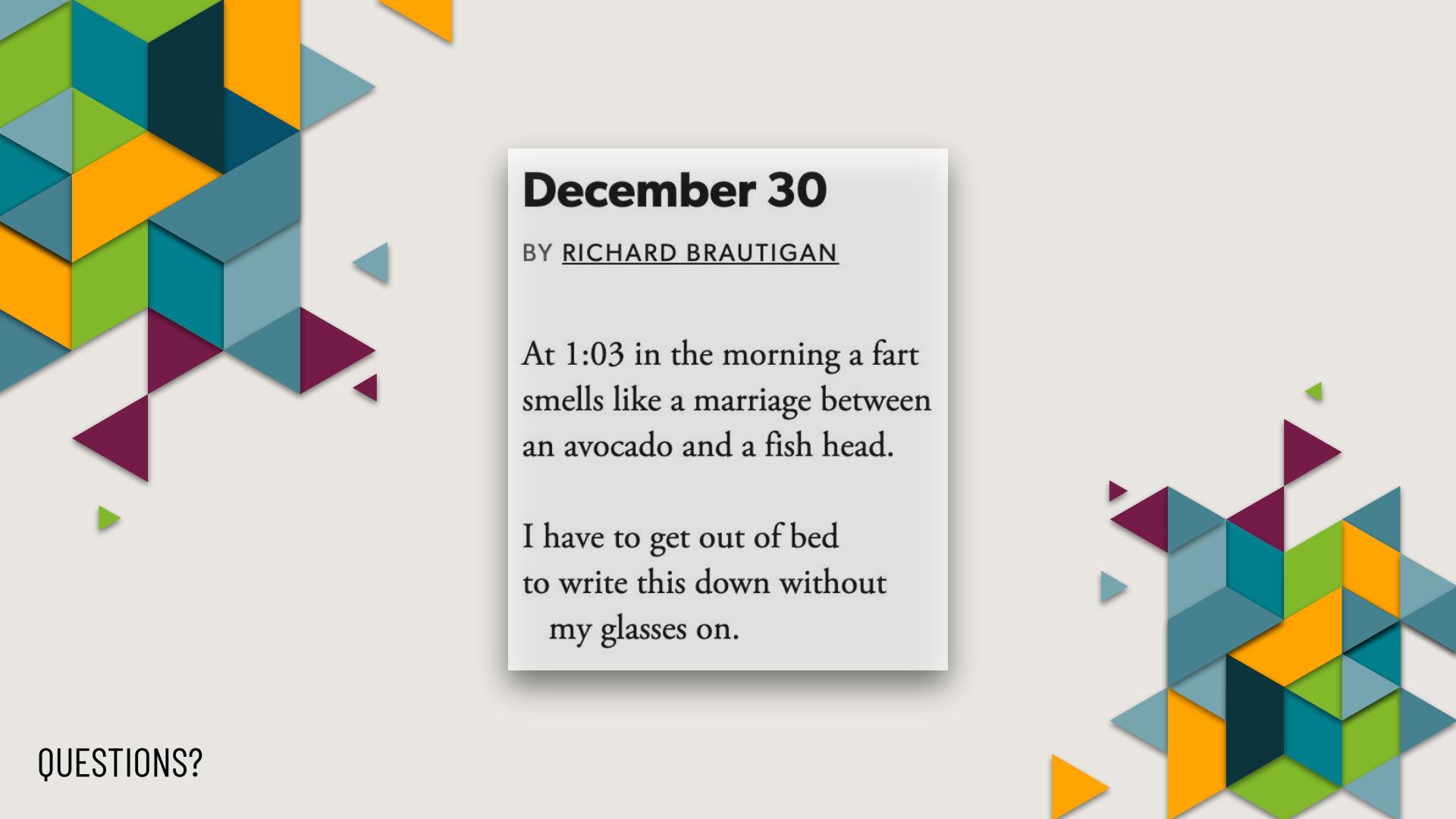
number of end rhymes, length of line,
syllables per word, number of lines,
polarity score, sentiment score

crazy, among, nobody, form, look
length of line, dream, small, cry, star



TAKEAWAYS

- Pre-1900 poetry is easily recognized
 - End rhymes and length of line
 - Avant-garde is its polar opposite
 - Form is important for most prediction models
 - SVM relies mostly on vocabulary
 - Poetry is rarely neutral (and generally positive)
- 



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?