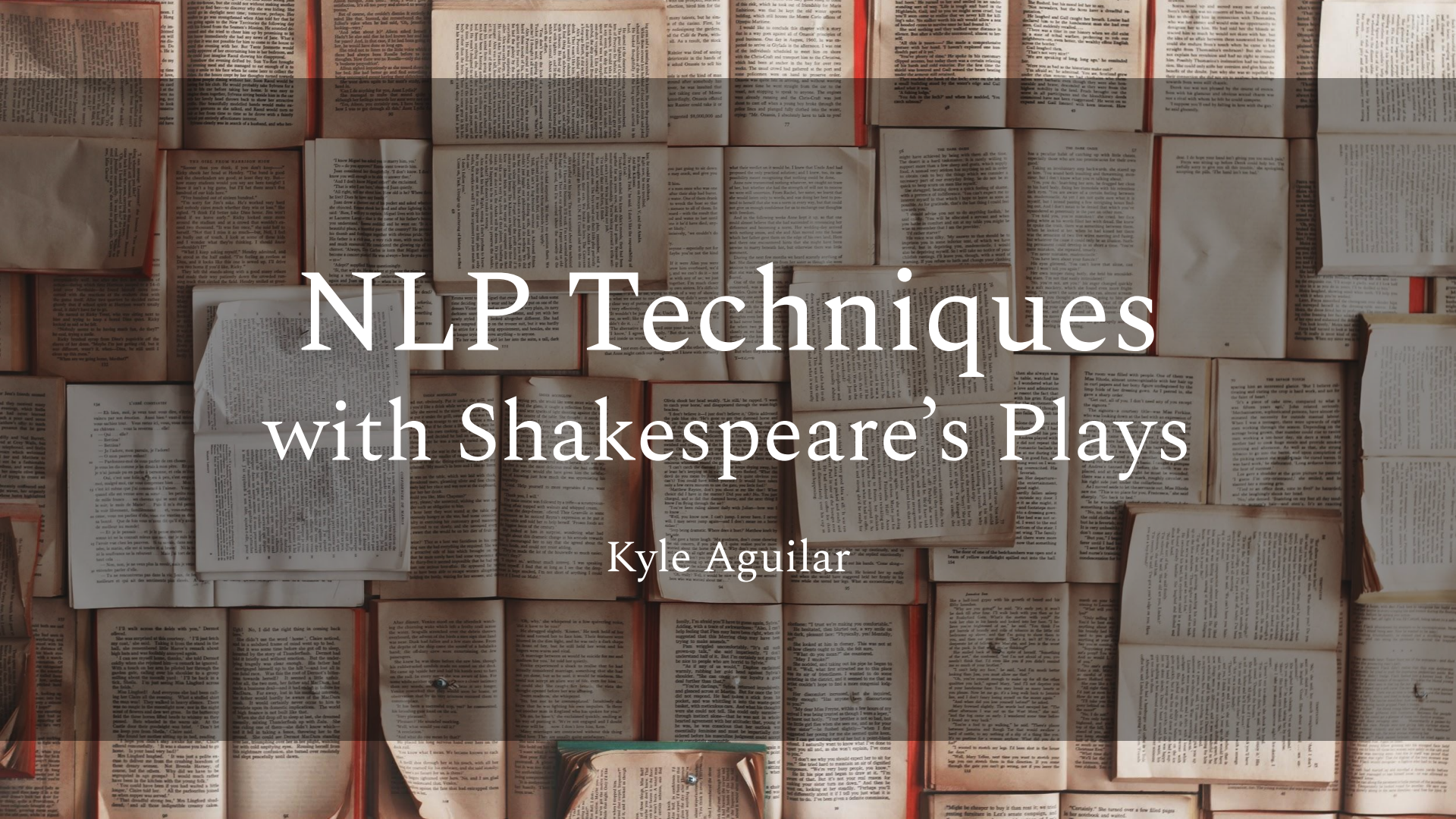
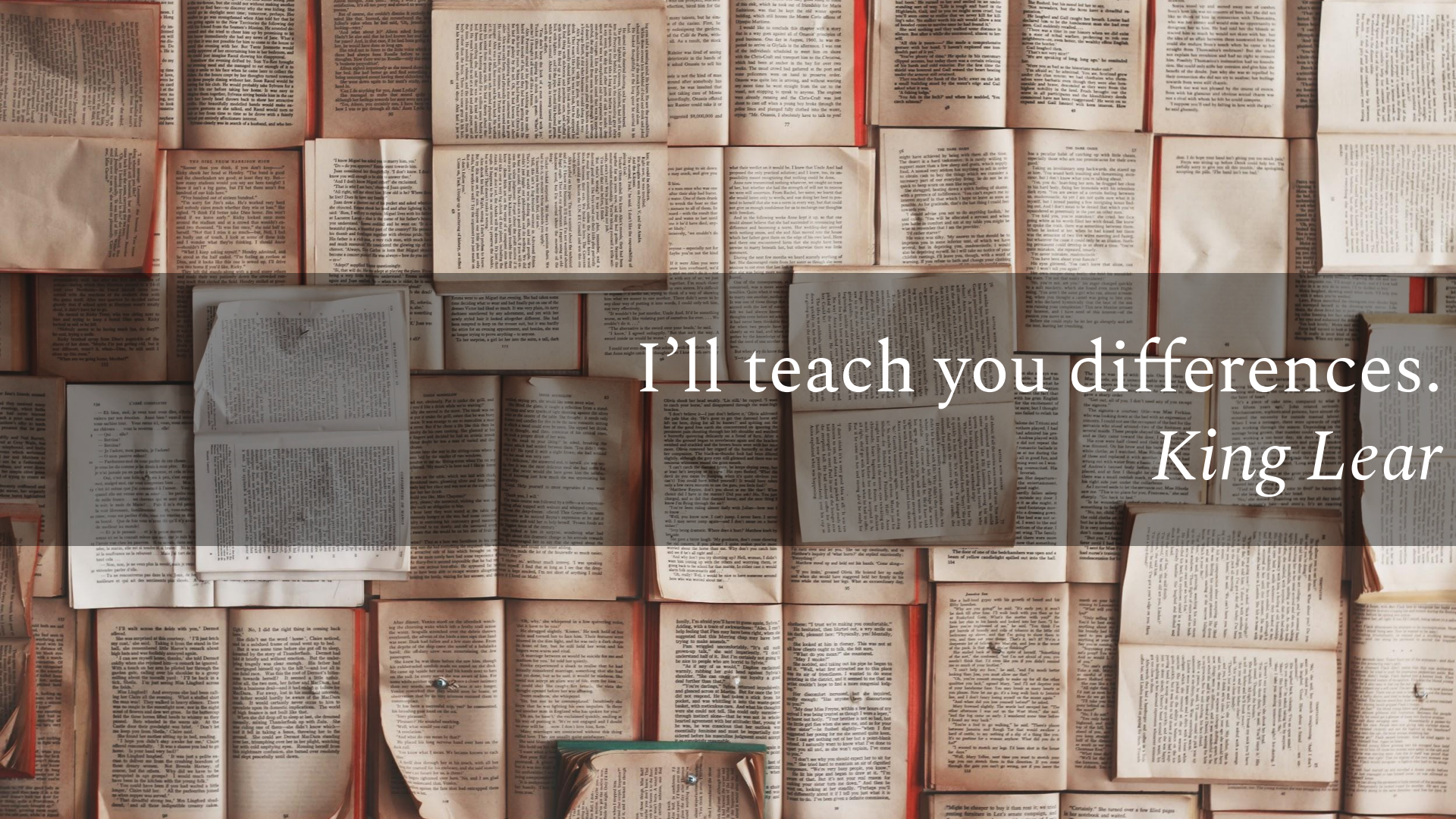
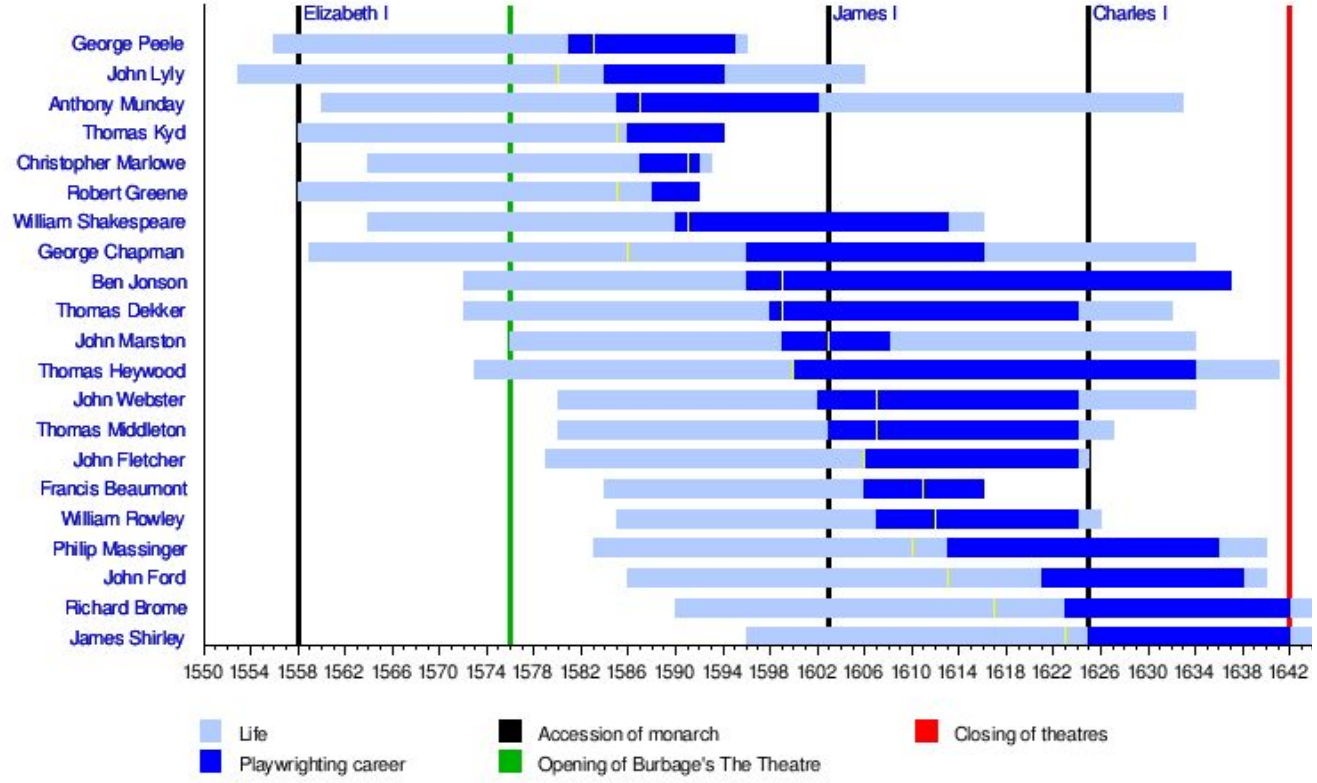


NLP Techniques
with Shakespeare's Plays

Kyle Aguilar


I'll teach you differences.
King Lear

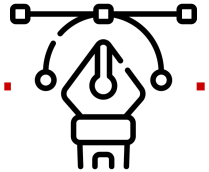
THE SHAKESPEARE AUTHORSHIP QUESTION



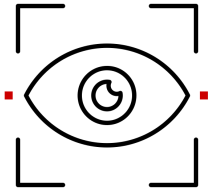
https://en.wikipedia.org/wiki/English_Renaissance_theatre

DATA PROCESS

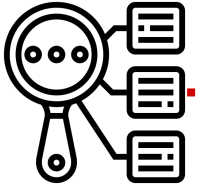
COLLECT PLAY TEXTS



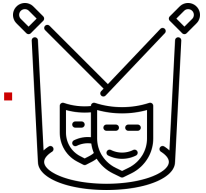
CLEAN / TRANSFORM



VISUALIZE TEXTS & CLASSES



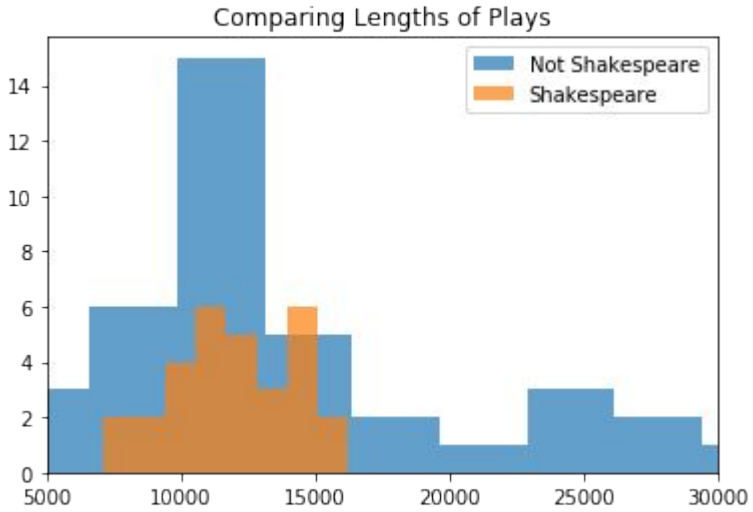
IMPLEMENT MODELS



EXPLORE TOPIC MODELING

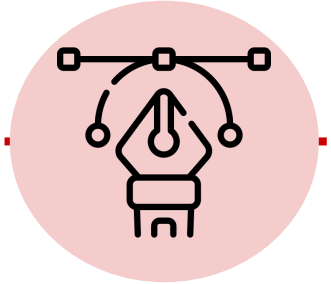
PLAY DATA

- [x] 39 Shakespeare Plays
- [x] 50 Non-Shakespeare Plays
- [x] Written from 1580s to 1630s
- [ ] Certainty about Authorship
- [ ] Equal Scholarship/Readership
- [ ] Consistency in Editing

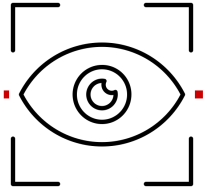


DATA PROCESS

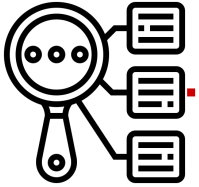
COLLECT PLAY TEXTS



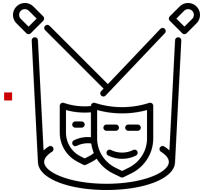
CLEAN / TRANSFORM



VISUALIZE TEXTS & CLASSES



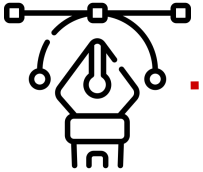
IMPLEMENT MODELS



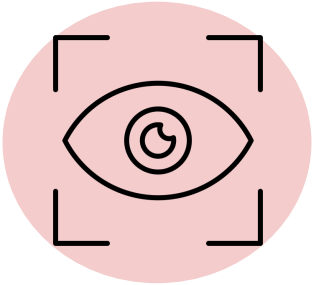
EXPLORE TOPIC MODELING

DATA PROCESS

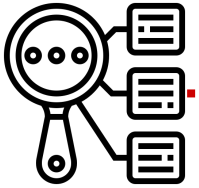
COLLECT PLAY TEXTS



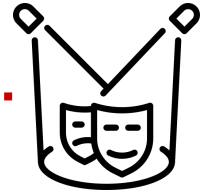
CLEAN / TRANSFORM



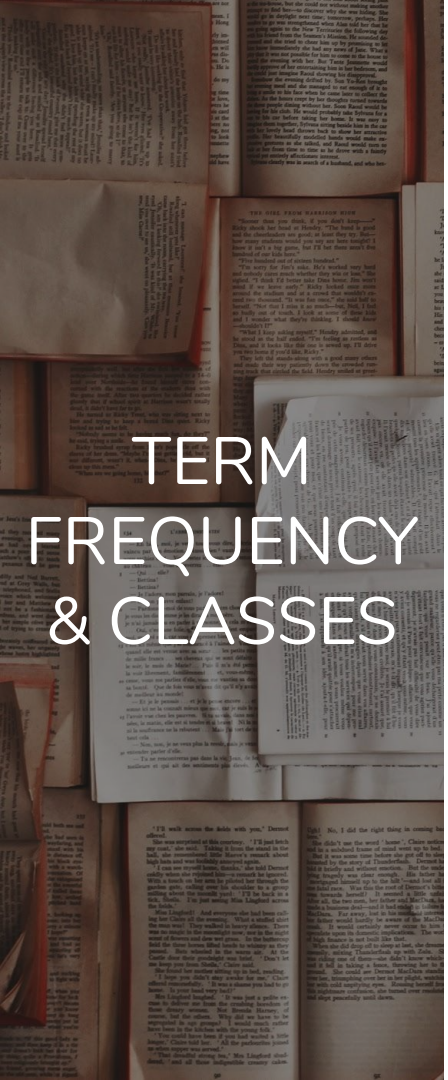
VISUALIZE TEXTS & CLASSES



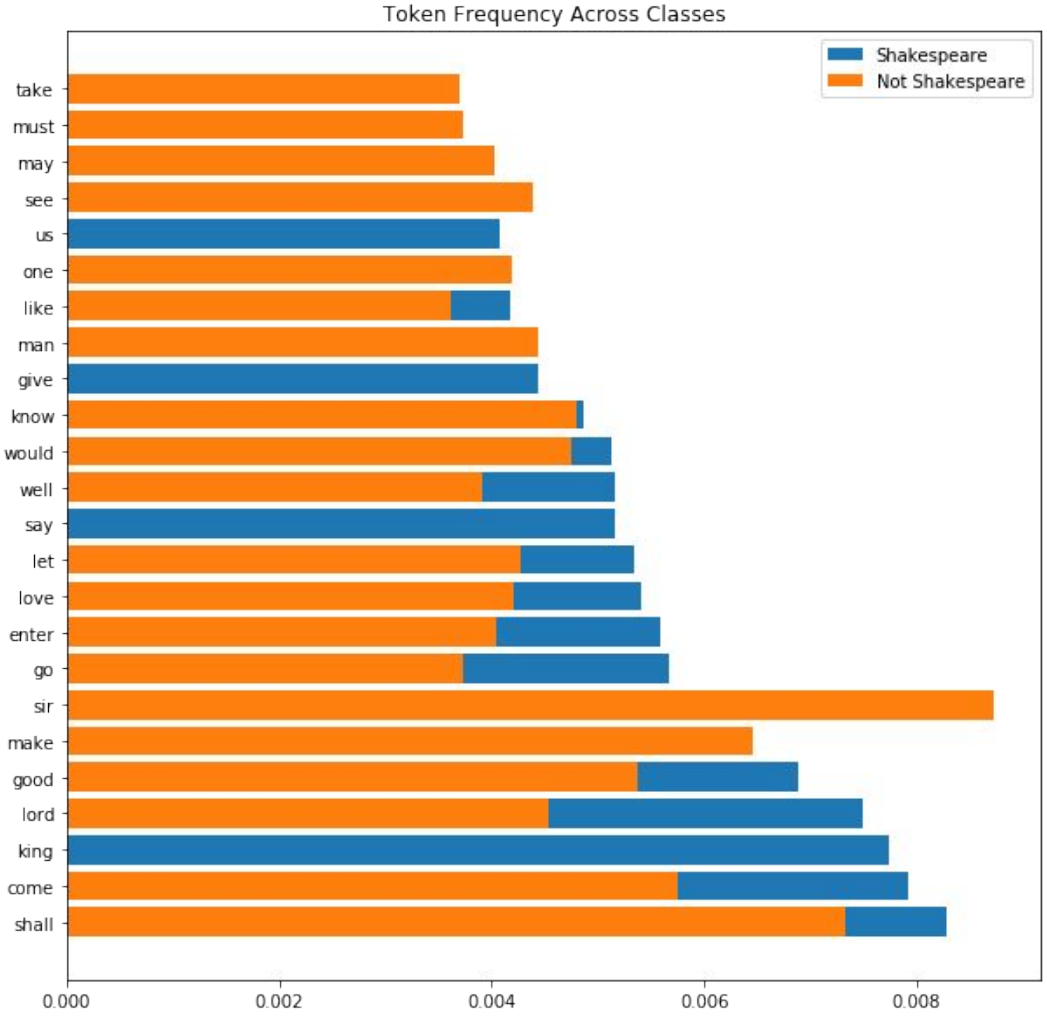
IMPLEMENT MODELS



EXPLORE TOPIC MODELING



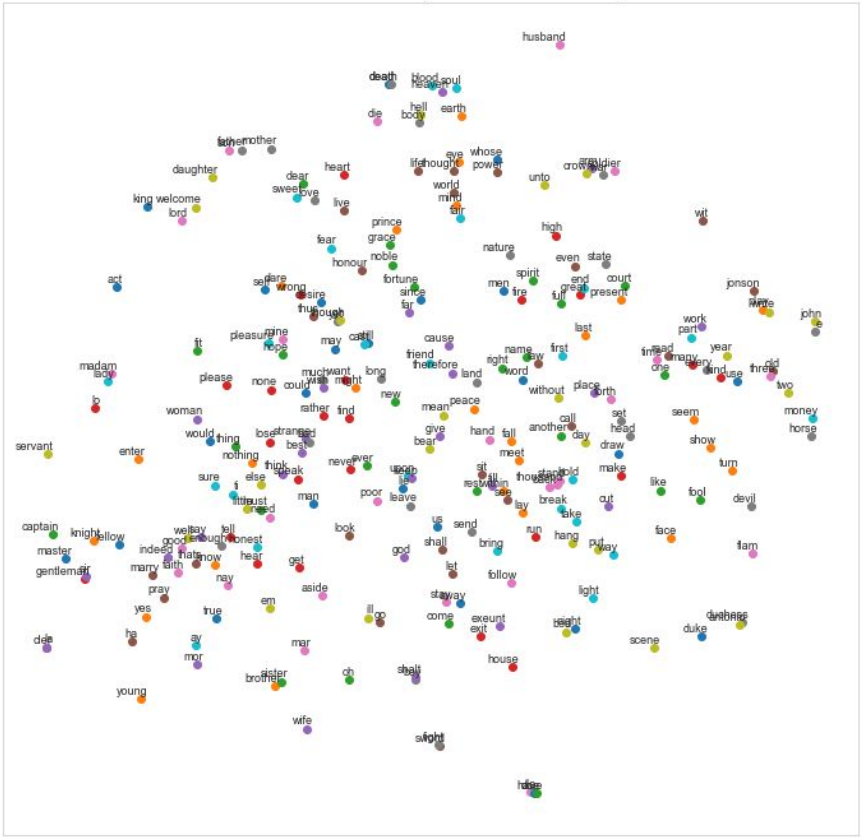
TERM FREQUENCY & CLASSES



CLASSIFYING TEXTS WITH MACHINE LEARNING

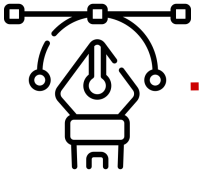
Most Common Shakespearean Words and Similarity

Most Common non-Shakespearean Words and Similarity

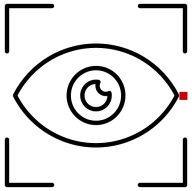


DATA PROCESS

COLLECT PLAY TEXTS

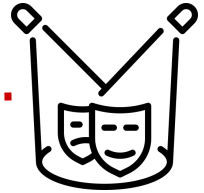


CLEAN / TRANSFORM



VISUALIZE TEXTS & CLASSES

IMPLEMENT MODELS



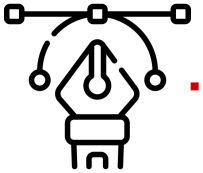
EXPLORE TOPIC MODELING

CLASSIFYING TEXTS WITH MACHINE LEARNING

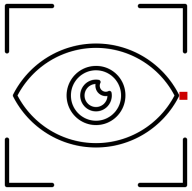
	Model_Name	F1_Score	Accuracy_Score	Precision_Score	Recall_Score
0	RF_TF-IDF	0.9412	0.9444	1	0.8889
1	RF_Dim-Reduc	0	0.5	0	0
2	RF_ngrams	0.6154	0.7222	1	0.4444
3	MNB_TF-IDF	0.6154	0.7222	1	0.4444
4	CNB_TF-IDF	0.6154	0.7222	1	0.4444
5	SVM_Dim-Red	0.3636	0.6111	1	0.2222
6	SVM_bigrams	0.5	0.6667	1	0.3333
7	XGB_TF-IDF	1	1	1	1
8	XGB_seq	0	0.5	0	0
9	RNN	1	1	1	1

DATA PROCESS

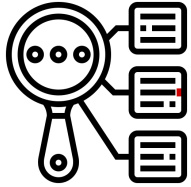
COLLECT PLAY TEXTS



CLEAN / TRANSFORM



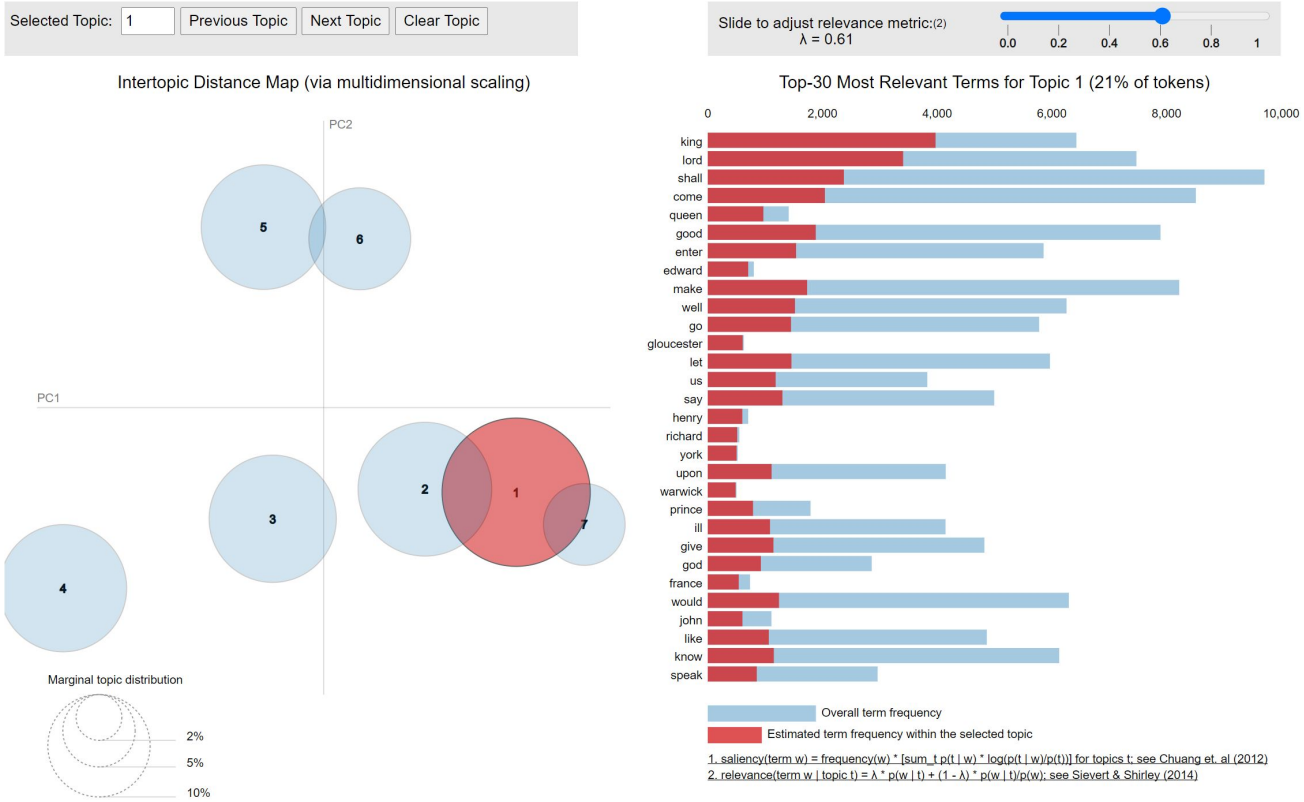
VISUALIZE TEXTS & CLASSES



IMPLEMENT MODELS

EXPLORE TOPIC MODELING

EXPLORING TEXTS WITH TOPIC MODELING

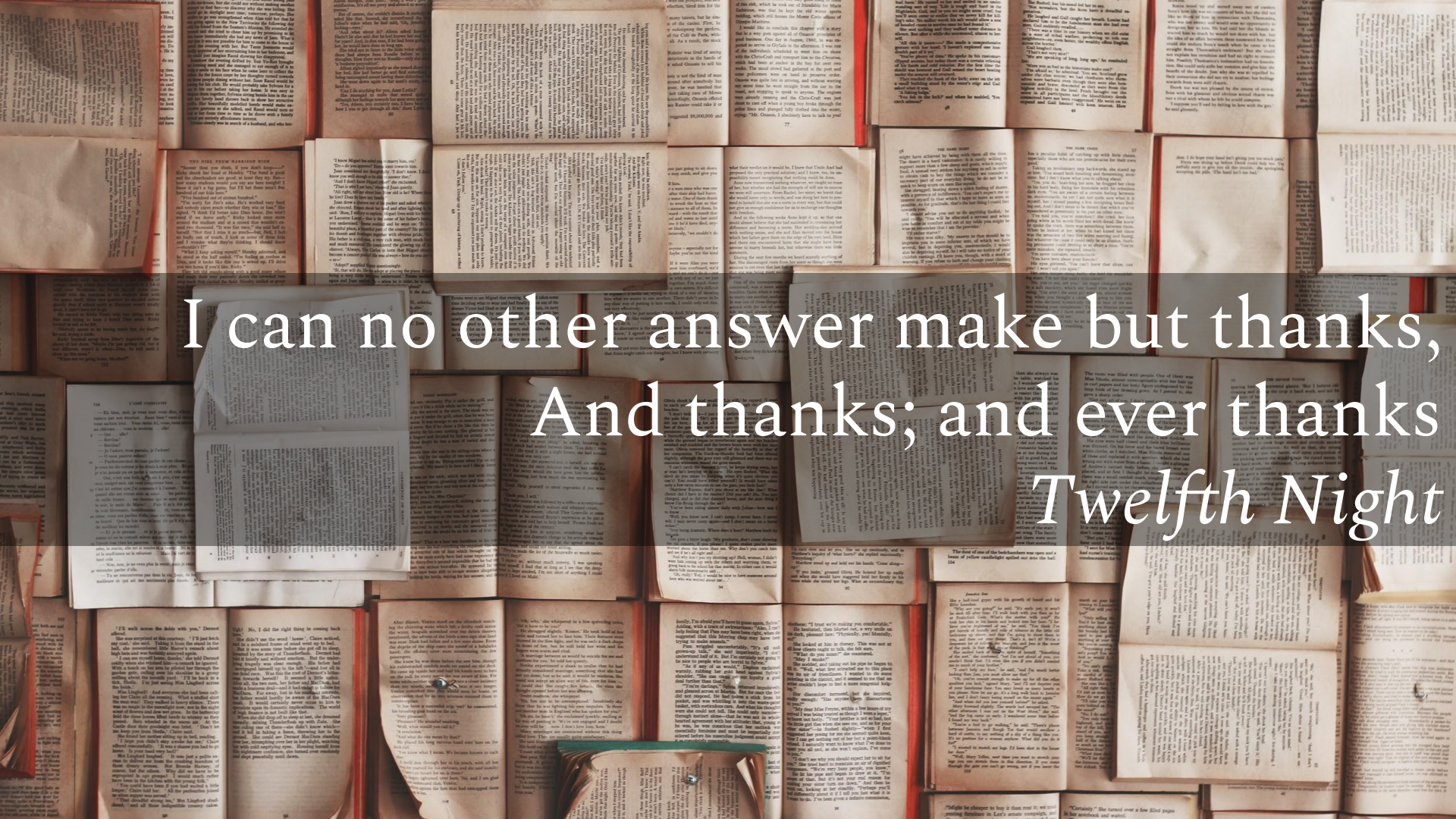


CONCLUSION

-

NEXT STEPS

- INCORPORATION OF SONNETS
- WEIGHTS/HYPERPARAMETER TUNING
- EXPAND NON-SHAKESPEARE DATASET
- BUILD RECOMMENDATION SYSTEM


I can no other answer make but thanks,
And thanks; and ever thanks
Twelfth Night