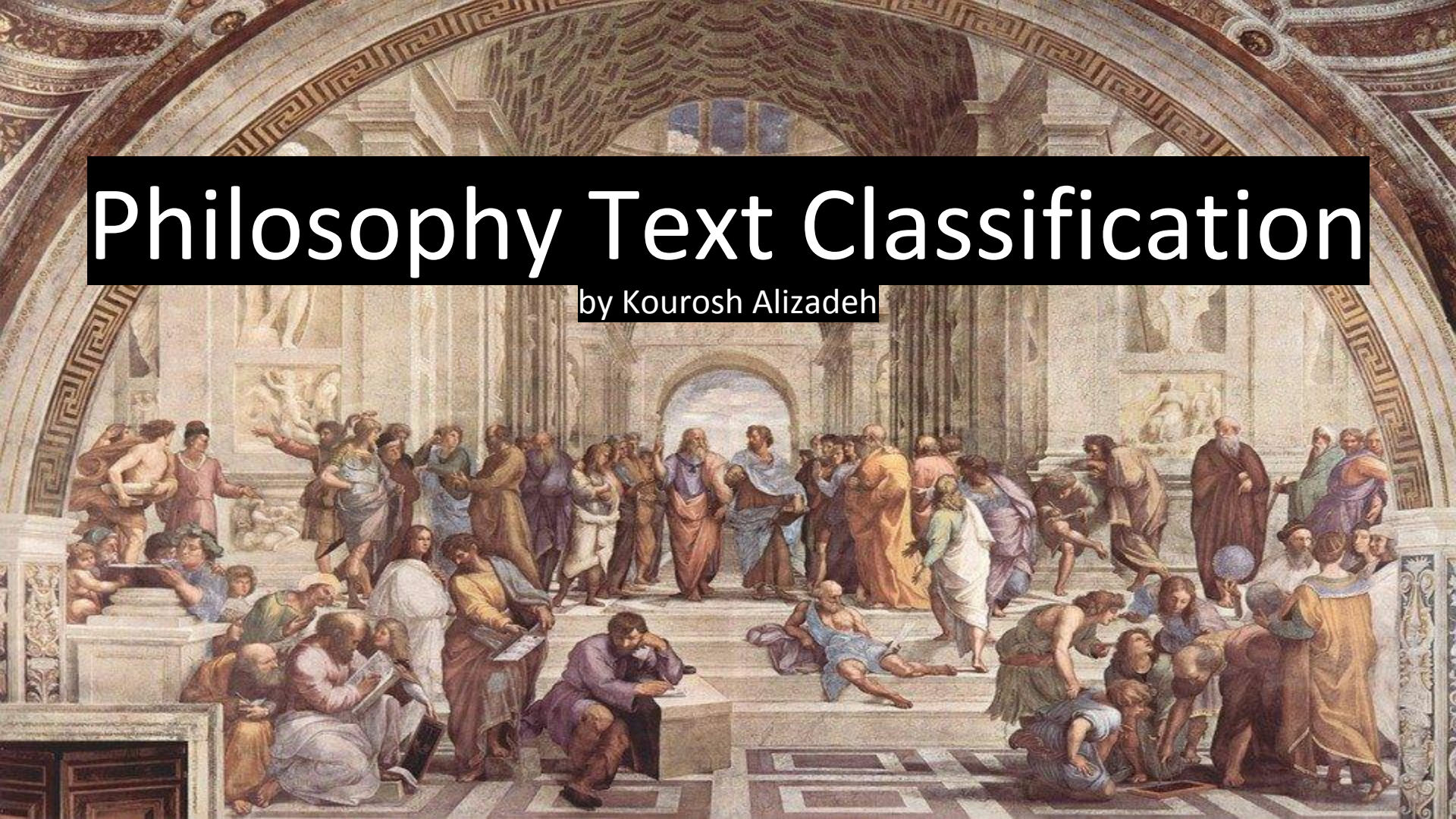# Philosophy Text Classification

by Kourosh Alizadeh

Why Classify Philosophical Texts?

- A philosophy represents a rationally developed worldview

- If we can classify texts based on their philosophical schools, we can classify sets of rationally organized perspectives on life

- This can be used in a purely academic way to isolate key features of a thinker's thought or notice differences hitherto unexplored

- If we expand this to the general populace, it means we can take user-generated text (tweets, FB posts) and identify a general worldview
  - this could be used in marketing, political campaigning, or sociological research

Word Clouds for Each School of Thought

Plato Word Cloud
think people me others things man friend say seem see take name right one could good SOCRATES ATHENIAN law THEAETETUS part god case kind tell come soul know like time

Aristotle Word Cloud
reason motion another say things movement part water belong animal possible body either one come kind number form case man good time like others nature place cause

Rationalism Word Cloud
things one sense cause truth good power nothing nature bodies think God idea true reason say motion certain order man soul know without men see love body object part

Empiricism Word Cloud
object men part sense man reason idea thought time knowledge certain passion principle think cause without another action nature nothing name mind things though power body word relation matter

German Idealism Word Cloud
object determination relation time unity existence idea appearance universal reason one thought principle consciousness concept self end understanding cognition content general nature form law nothing moment world means

Phenomenology Word Cloud
essence consciousness things sense knowledge experience possible perception form always one subject time object truth present existence thought fact say world body meaning within science become work beings question

Analytic Word Cloud
sentence world true see term fact one time object say sense use meaning given could belief name p proposition question seem language think theory word statement know case

Continental Word Cloud
point become relation idea knowledge order time object one within question reason element language madness form thought man truth sign could nature representation difference sense world word desire possible
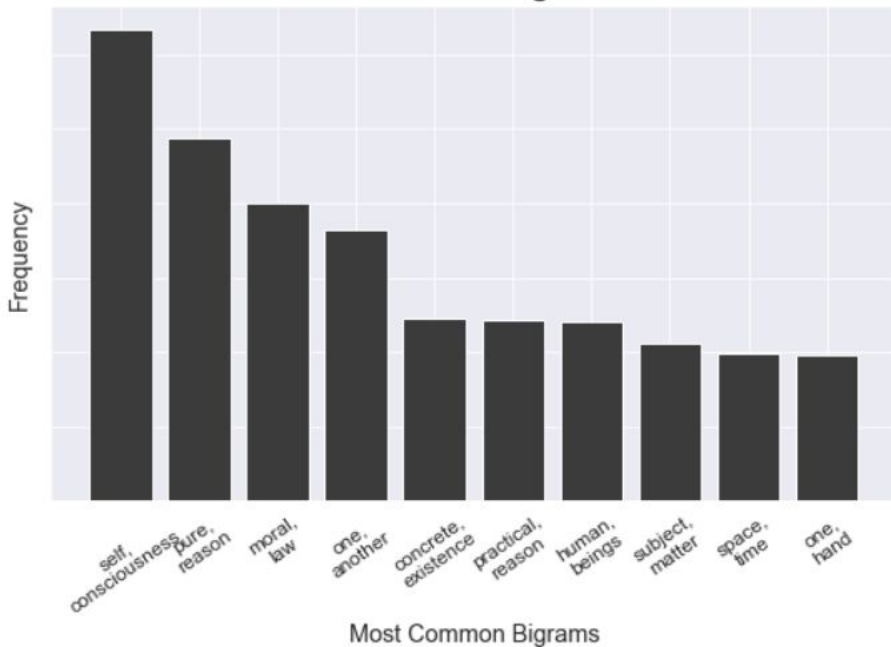
left page intentionally this blank
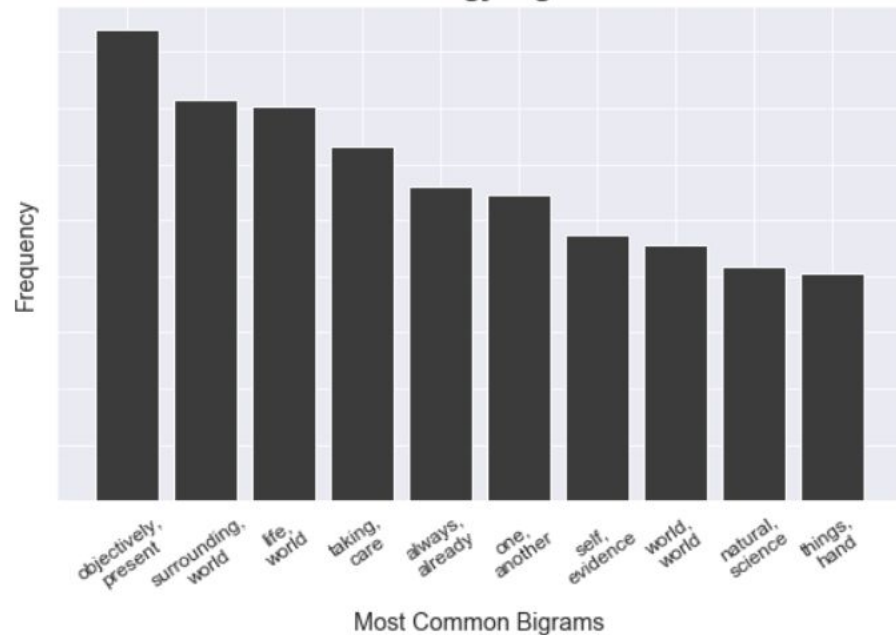
Overall Word Frequency
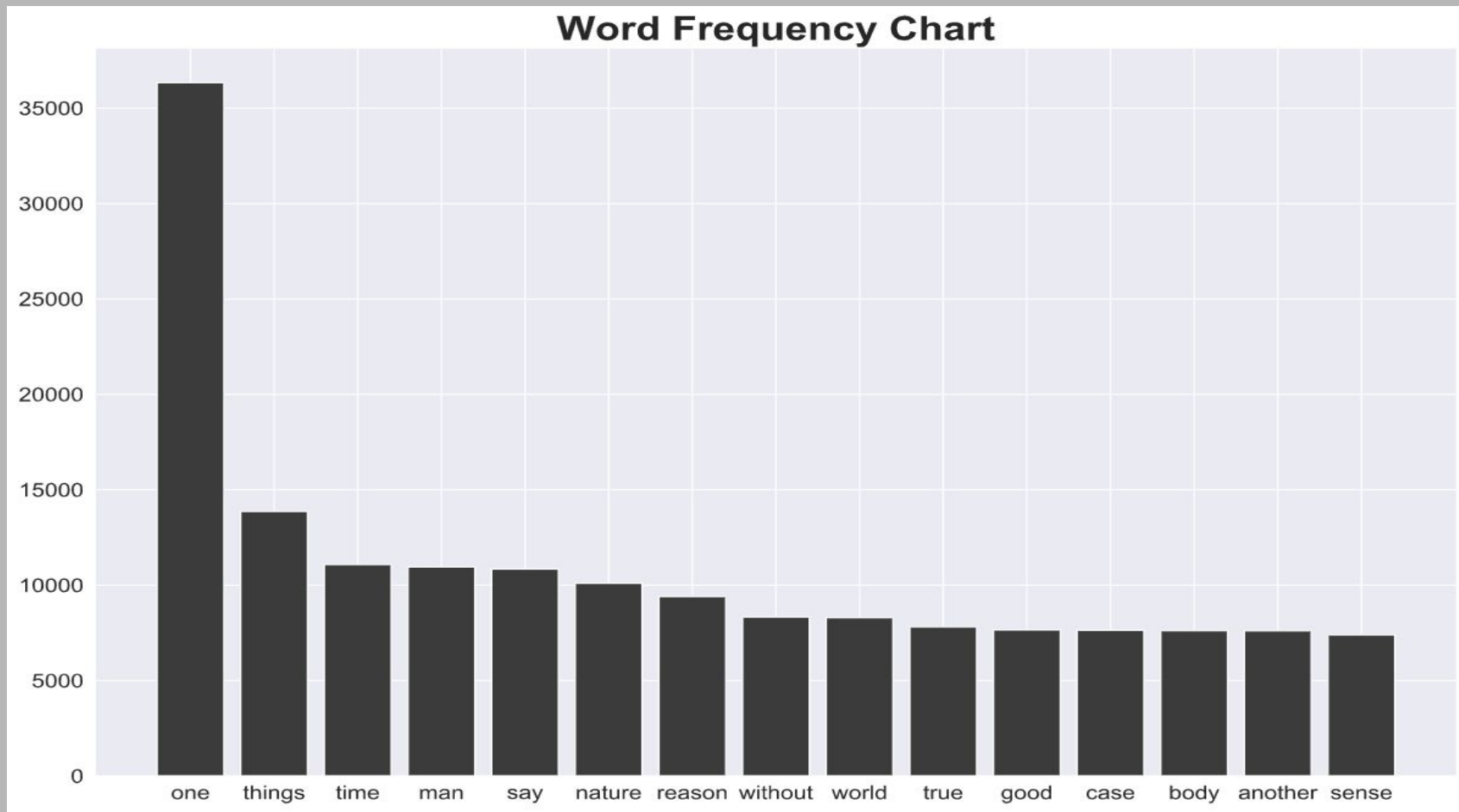
A Few Illuminating Bigrams

- self consciousness is a key concept in German Idealism
- the idea of objective presence and the phrase 'always-already' are common in Phenomenology
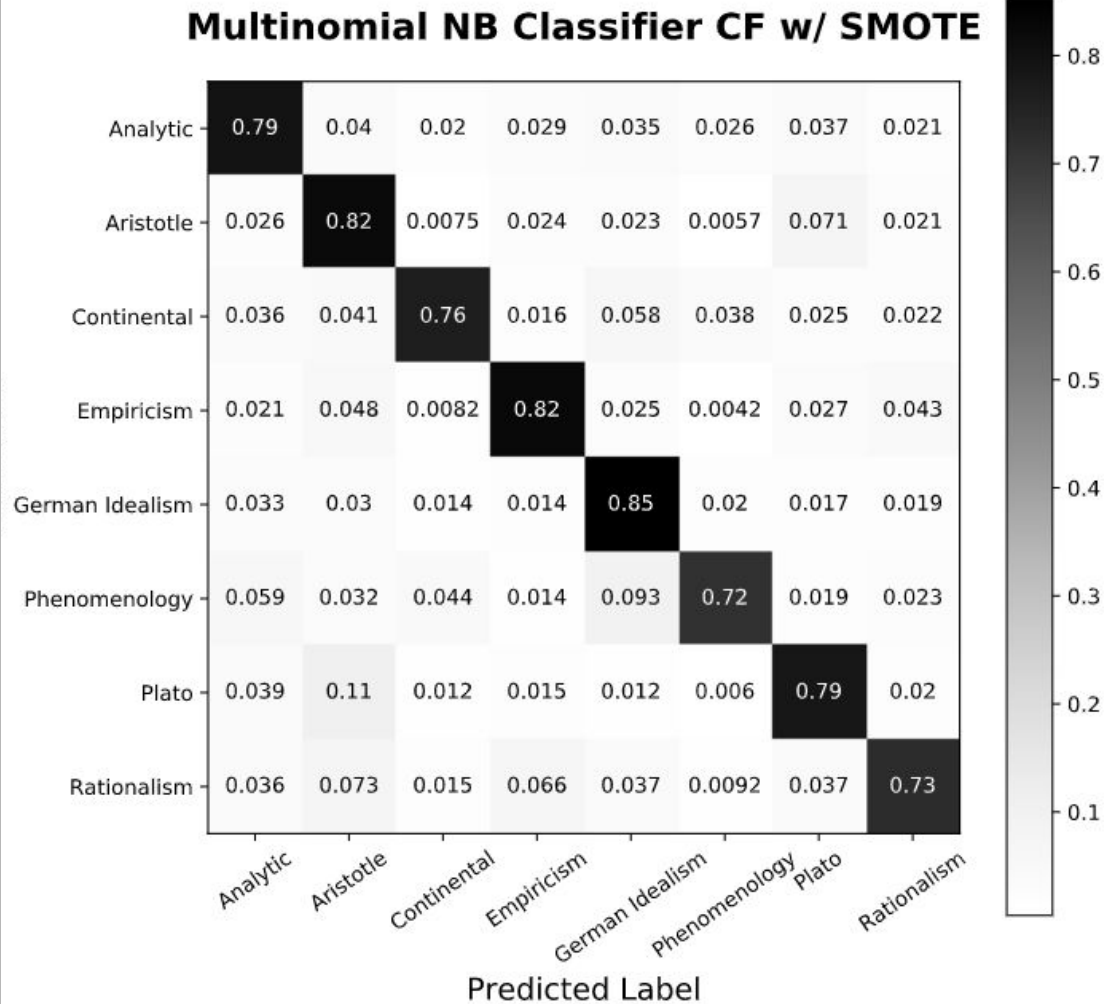

German Idealism Bigram Chart


Phenomenology Bigram Chart

Bayesian Model

- Better performance than Random Forest

- 79% accuracy rate

- Aristotle, Empiricism, and German Idealism have strong recall rates.

Multinomial NB Classifier CF w/ SMOTE

	Analytic	Aristotle	Continental	Empiricism	German Idealism	Phenomenology	Plato	Rationalism
Analytic	0.79	0.04	0.02	0.029	0.035	0.026	0.037	0.021
Aristotle	0.026	0.82	0.0075	0.024	0.023	0.0057	0.071	0.021
Continental	0.036	0.041	0.76	0.016	0.058	0.038	0.025	0.022
Empiricism	0.021	0.048	0.0082	0.82	0.025	0.0042	0.027	0.043
German Idealism	0.033	0.03	0.014	0.014	0.85	0.02	0.017	0.019
Phenomenology	0.059	0.032	0.044	0.014	0.093	0.72	0.019	0.023
Plato	0.039	0.11	0.012	0.015	0.012	0.006	0.79	0.02
Rationalism	0.036	0.073	0.015	0.066	0.037	0.0092	0.037	0.73

Predicted Label

Random Forest Model Feature Importances

- RF was less accurate but gave us feature importances.

- 'god' is a major topic in some schools but not others,
- 'socrates' is much more common in Plato than elsewhere.

Feature Importances RF Classifier
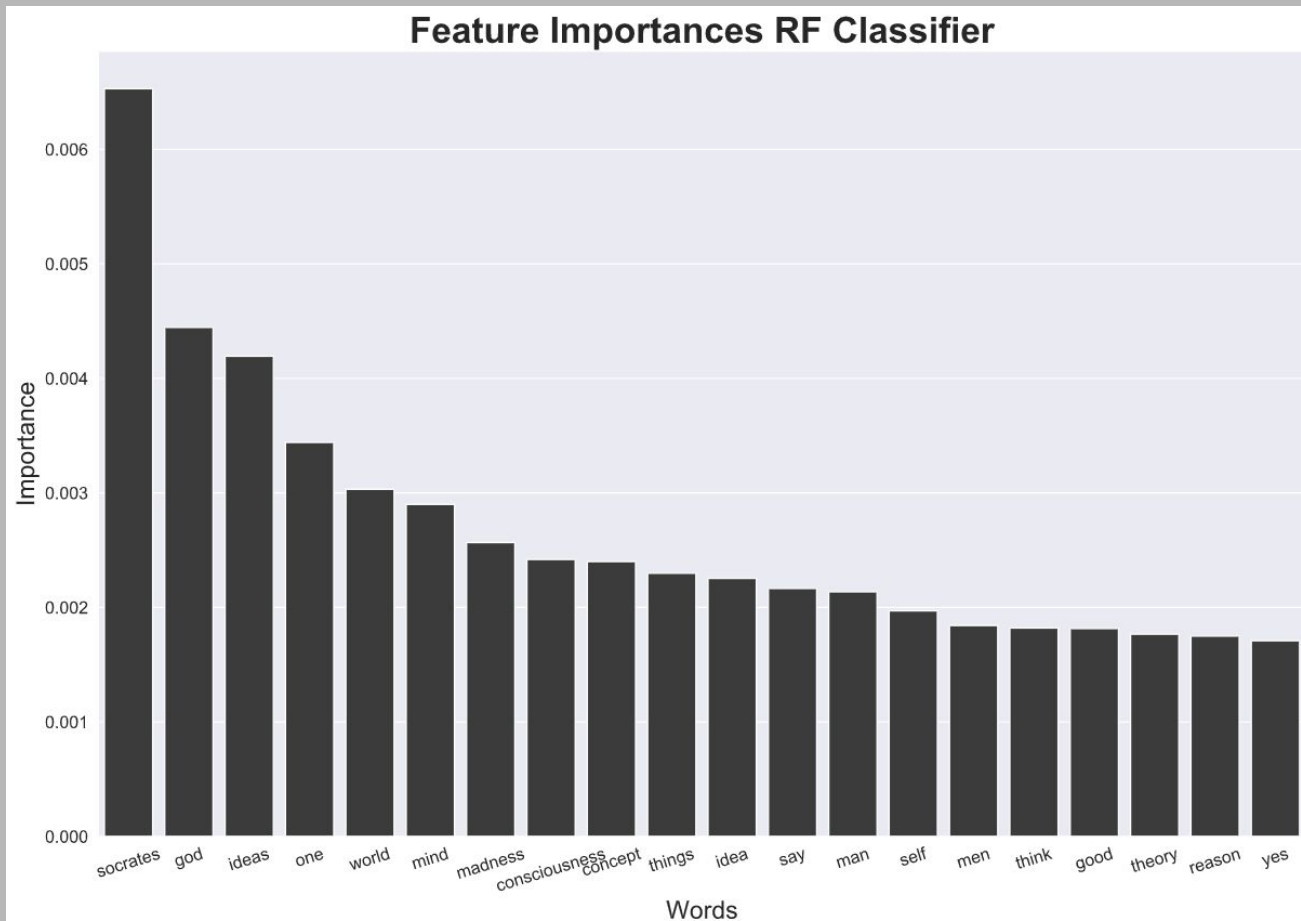
Key Takeaways & Recommendations

- The model can be used to identify schools with about 80% accuracy.

- Analyze a client's text to identify ideology and preferences at a deep level
    - then tailor communications and persuasions to that ideology
    - for best results use the client's key terms and do not contradict the beliefs of their schools

- Key classificatory terms include 'mind,' 'think,' 'god,' and 'madness.'
    - so look for client's attitude toward these topics.

- For academic research, look at word/bigram frequency to get an idea of how to orient yourself in studying a school

- Proof-of-concept operationalization in a notebook that classifies user text

Future Work

- Clean the data more thoroughly, especially for typos and word fusing that was common in the data (e.g., 'aconcept' for 'a concept').

- Develop more effective neural networks.

- Take the user-input notebook and set it up in a webapp.

- As an extreme goal, it would be nice to to build a generative model that could create text in the style of a philosopher or paraphrase given text in the style or worldview of a given school.
  - this would enable us to generate text to fit a client's worldview after identifying it

Thank you