

# Classifying Philosophical Texts by School of Thought

Brianna Klabunde



# Can we create a model that accurately classifies philosophical texts by their school of thought?

**Organizing Knowledge:** Automatically classifying texts helps researchers, students, and readers navigate large philosophical corpora and quickly find works related to a specific school of thought (e.g., Stoicism, Phenomenology).

**Supporting Analysis:** Enables quantitative studies—such as tracking how ideas evolve across time or comparing linguistic and conceptual patterns between traditions.

# About the data

## → 5 schools

Stoicism, Feminism, German Idealism, Phenomenology,  
Analytic

## → 5000 samples

1000 rows for each school

## → Lemmatized strings

Extended contractions > lower case letters > removed  
stop words > lemmatized, but kept pronouns

→ Alizadeh, K. (2019). *History of Philosophy* [Data set]. Kaggle.



1

Brief overview  
of each school

2

Text example  
from the  
dataset

3

What makes  
each school  
distinct?



# Stoicism

**“If reason is general, then is that reason also, which prescribeth what is to be done and what not, common unto all.”**

- Marcus Aurelius, *Meditations*

- **Focus:** An ethical philosophy concerned with self-control, rationality, and living in accordance with nature.
- **Language & style:** Formal, old-fashioned language using imperatives, and ethical maxims rather than personal examples.

1

stoicism

This word cloud illustrates the frequency of specific words in the King James Version of the Bible. The size of each word corresponds to its relative frequency in the text. The words are color-coded into three main groups:

- Green words:** include common, good, well, make, part, hath, live, time, world, thou, art, unto, thyself, possible, whether, rational, hinder, go, come, universe, yet, natural, whole, unto, thee, word, reasonable, seek, substance.
- Blue words:** include doth, one, remember, right, free, pleasure, require, consider, whose, several, general, thy power, matter, neither, happen unto, proper, upon, thou must, life, may, kind, much, change, work, still, put, avoid, look, certain, keep, apply, reasonable, seek, substance.
- Purple words:** include especially, among, preserve, love, thou mayest, object, appoint, particular, reasonable creature, thou hast, another, thou dost, doth happen, mean, power, able, receive, grief, first, though, might, say, death, think, likewise, altogether, praise, toward, occasion, word, reasonable, seek, substance.

The background features a faint watermark of the text "THE KING JAMES BIBLE".

# Feminism

**“Isabelle Eberhardt, who dressed as a man and traversed the Sahara on horseback, felt no less diminished when she gave herself to some vigorous sharpshooter.”**

- Simone De Beauvoir, *The Second Sex*

- **Focus:** Examines gender, power, and social inequality
- **Language & style:** First-person, significance of pronouns, and concrete examples-- often discusses politics, society, and relationships

2

# feminism

A word cloud centered around the word "feminism". The words are arranged in a circular pattern, with the size of each word indicating its frequency or importance. The colors of the words vary, creating a vibrant and dynamic visual effect.

The most prominent words in the center include:

- feminism** (large, black)
- woman** (large, purple)
- man** (large, green)
- love** (large, blue)
- child** (large, purple)
- work** (large, green)
- world** (large, green)
- husband** (large, blue)
- society** (large, blue)
- give** (large, green)
- make** (large, green)
- time** (large, green)
- right** (large, blue)
- WANT** (large, purple)
- long** (large, blue)
- take** (large, blue)
- case** (large, blue)
- thus** (large, blue)
- mother** (large, green)
- first** (large, green)
- even** (large, blue)
- see** (large, blue)
- power** (large, yellow)
- always** (large, yellow)
- accept** (large, yellow)
- situation** (large, blue)
- sense** (large, blue)
- whose** (large, blue)
- discover** (large, blue)
- nothing** (large, blue)

Surrounding these central words are numerous smaller words, many of which are repeated multiple times, creating a dense and complex network of associations:

- earth, mystery, social, instead, affection, reach, law, bear, home, reason, pleasure, bring, without, girl, less, explain, hold, respect, two, yet, old, try, free, passion, keep, show, duty, father, dream, reality, tell, sister, possess, relation, demand, human, education, fact, lead, let, good, role, daughter, write, wife, form, object, experience, leave, movement, black, woman, family, prove, call, male, young, death, mean, believe, truth, look, play, true, go, begin, act, sex, away, nature, master, live, necessary, force, appear, many, must, eye, may, face, sexual, slave, might, course, value, every, among, create, hand, remain, beauty, end, marriage, much, seem, come, heart, find, particular, struggle, carry, mind, black, arm, speak, self, far, marry, certain, individual, never, year, difficult, think, God, kill, order, flash, part, thing, de, take, new, need, well, word, within, lover, way, 8

# German Idealism

“Itself become the basis of the existence of objects and insofar as reason, through this cognition, has causality in a rational being, as a pure reason that can be regarded as a power directly determining the will.”

- Kant, *Critique of Practical Reason*

- **Focus:** Explores how reason, consciousness, and the mind shape reality; seeks to understand the connection between thought, experience, and objects.
- **Language & style:** Dense, abstract, and formal; frequent use of long, complex sentences, technical philosophical terms

3

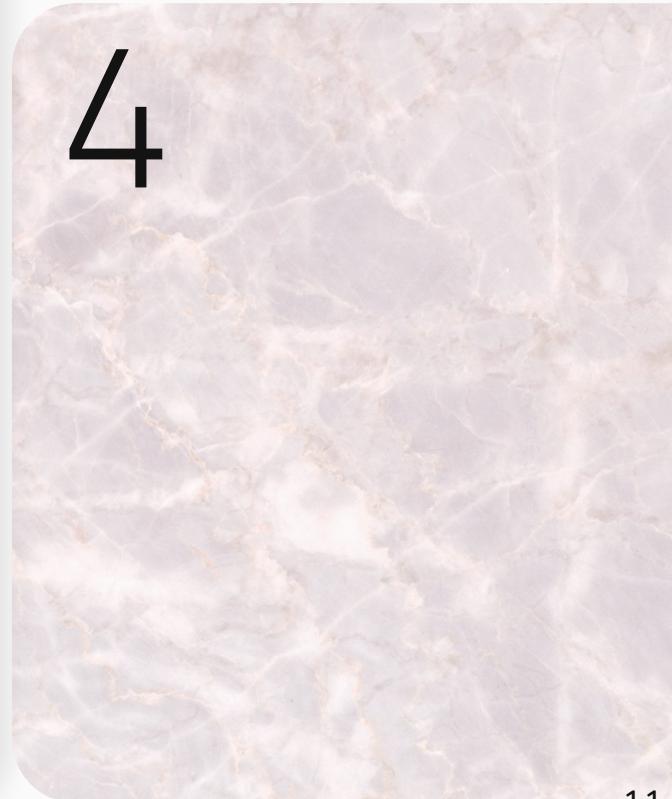
## german\_idealism

# Phenomenology

“There is therefore reason for holding a priori that all the senses are spatial, and the question which is the one which presents us with space must be considered unintelligible if we reflect on just what a sense is.”

- Merleau-Ponty, *The Phenomenology of Perception*

- **Focus:** How we experience and understand the world.  
What is the nature of conscious experience?
- **Language & style:** Dense, reflective, often first-person;  
describes experience and trust in knowledge



phenomenology

must science say knowledge structure character without  
nanner reality point disclosure necessarily  
future already presence become relation  
side risk indicate pure life world idea situation  
phenomenon reason rather presentation  
language first body nothing new concept precisely  
phenomenal temporal human word natural historical  
object never arise may place set fact come psychology  
expression stand thinking whole possibility  
man perceive power method order something history  
case belong change concern lead theory part  
Dasein God basis moment  
long see thought part  
work meaning Da sein interpretation  
work meaning Da sein interpretation  
way subject temporality analysis  
way subject temporality analysis  
think take toward back take care validity  
make know existential significance indeed  
know life space soul  
self objective look certain kind  
call bring remain essential ground immediately  
upon sphere field two direction  
consciousness even shape attitude colour essence within

# Analytic

**“His argument is valid, and his conclusion is correct, provided its (perhaps dubitable) premise is accepted.”**

- Kripke, *Philosophical Struggles*

- **Focus:** Emphasizes clarity, logic, and argumentation, often analyzing language, meaning, and scientific reasoning.
- **Language & style:** Clear, precise, and technical; uses short, logical sentences, formal definitions, and structured arguments

5

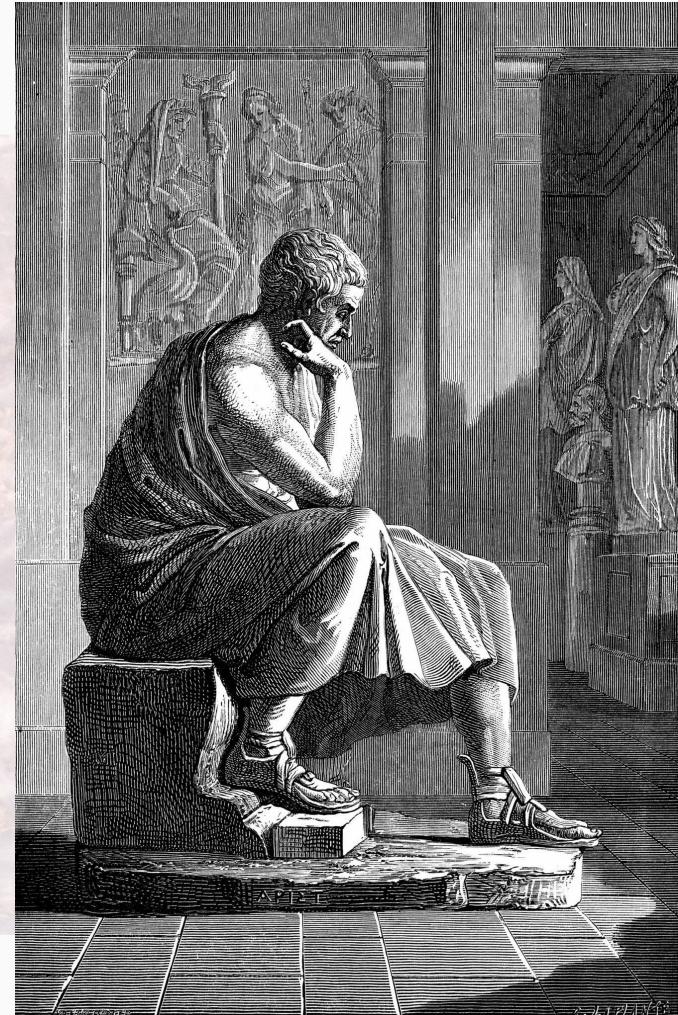
analytic

# Feature Engineering

Countvectorizer (lemmatized tokens)

TF-IDF (lemmatized tokens)

TF-IDF (ngram(2, 3))



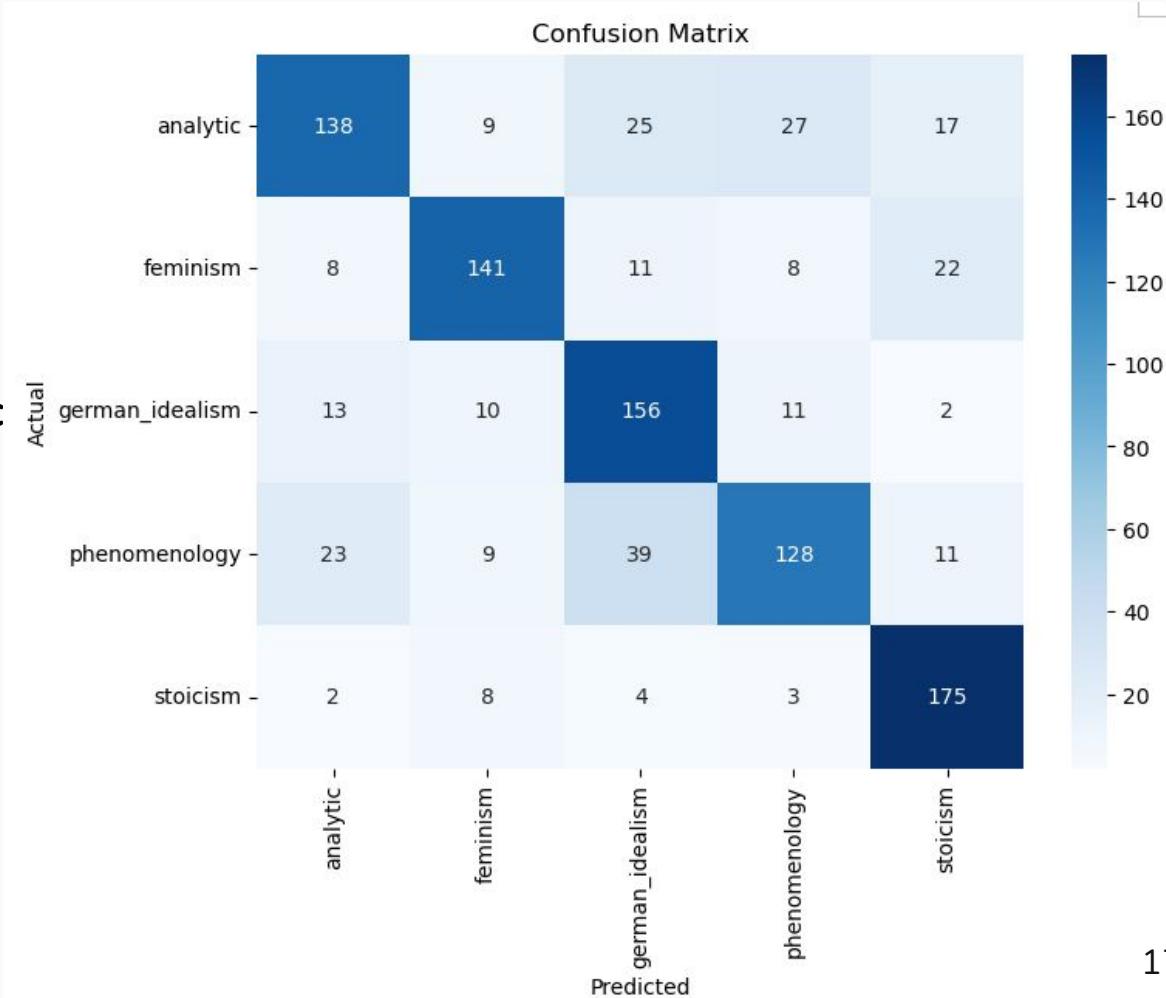
# Model performance by accuracy score

	Count Vectors	TF-IDF	N-Gram (2, 3)
Naive Bayes	73.7%	73.8%	41.8%
Logistic Regression	71.5%	72.9%	41.2%

- Random Forest and Gradient Boosting performance below 65% accuracy

# Evaluation

- **German Idealism** frequently overpredicted (especially for Phenomenology and Analytic)
- **Phenomenology and Analytic** are often confused, reflecting shared structures
- **Stoicism and Feminism** occasionally misclassified as each other, suggesting overlap



# F1 scores on each class

- Phenomenology, analytic, german-idealism had the lowest F1 scores
- German idealism only did well because it was over predicted (precision of 66%)
- These three classes were very aligned in style, clear from word clouds
- Stoicism and feminism were the most distinct

-Precision: of all predicted instances of this class, how many were correct?

-Recall: of all actual instances of this class, how many did the model find?

## Classification Report:

	precision	recall	f1-score	support
analytic	0.7500	0.6389	0.6900	216
feminism	0.7966	0.7421	0.7684	190
german_idealism	0.6638	0.8125	0.7307	192
phenomenology	0.7232	0.6095	0.6615	210
stoicism	0.7709	0.9115	0.8353	192
accuracy			0.7380	1000
macro avg	0.7409	0.7429	0.7372	1000
weighted avg	0.7407	0.7380	0.7346	1000

# Conclusion

- **Classification is possible:** Philosophical texts can be categorized by school of thought
- **Performance could improve:** Larger datasets, longer text samples, more rigorous cleaning, and tuning model parameters could likely yield more accurate results
- **Model comparison:** Naive Bayes and Logistic Regression performed well, holding up against more complex boosting and bagging methods
- **Feature comparison:** N-gram features performed poorly, while Count Vectorizer and TF-IDF gave similar results.



# Thank you

