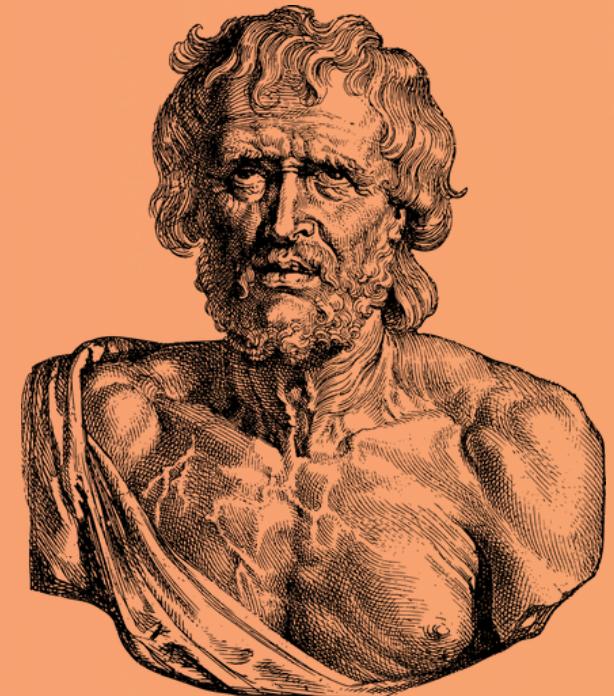


Reddit NLP

Classifying Philosophical Discourse r/Stoism and r/Buddhism



[source](#)



[source](#)

“

“The ultimate goal of the Stoic was apatheia, or peace of mind, which I think is akin to... the Buddhist goal of nirvana...”

Professor Massimo Pigliucci

Introduction

"The happiness of your life depends upon the quality of your thoughts."

Marcus Aurelius

"The mind is everything. What you think you become."

The Buddha

"The wise man accepts his pain, endures it, but does not add to it."

Marcus Aurelius

"Pain is certain, suffering is optional."

The Buddha

The Similarities

- Advocate for **introspection, observation, and meditation**.
 - Exalt the **present moment** as the path to contentment and truth.
 - Emphasize the **impermanence** of all things.
 - Offer practical, **secular philosophies** for living a better life.
 - Focus on **inner peace**, self-control, and personal responsibility.
 - Utilise **meditation** as a tool for reflection and growth.
 - View the **universe as interconnected** and teach that the universe is one.
 - Teach to **embrace death**.
-

Problem Statement

Many similarities have been drawn between the teachings of Buddhism and Stoicism. Given this, can the modern discourse between the two be accurately classified (with 90% or more accuracy) on unseen data?

If so, does this hold, even as technical terminology, philosophers' names and famous writings are excluded from the model?

Project Goals

BUILD A
CLASSIFICATION
MODEL WITH



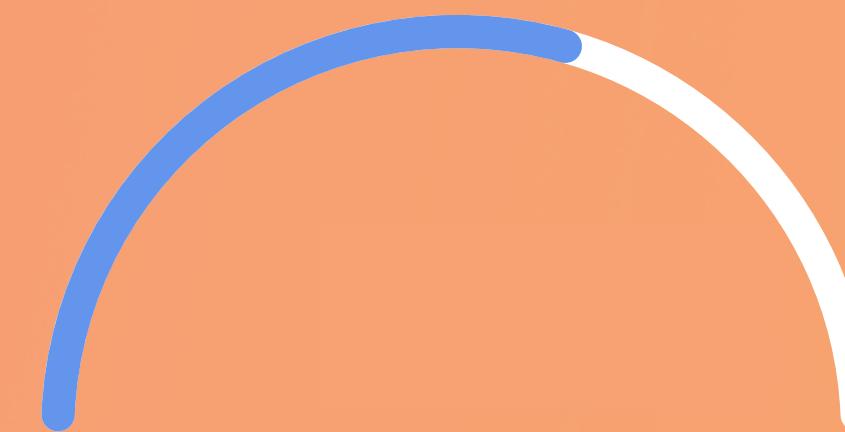
OR MORE ACCURACY
ON UNSEEN DATA

REMOVE TECHNICAL
TERMINOLOGY



TO SEE HOW THE
MODEL IS AFFECTED

REVEAL INSIGHTS
THROUGH
COMPARISON



BETWEEN SUBREDDIT
DISCOURSE

Cleaning & EDA

A Brief Overview



[source](#)

Baseline 51% the
buddhism subreddit
(negative class)

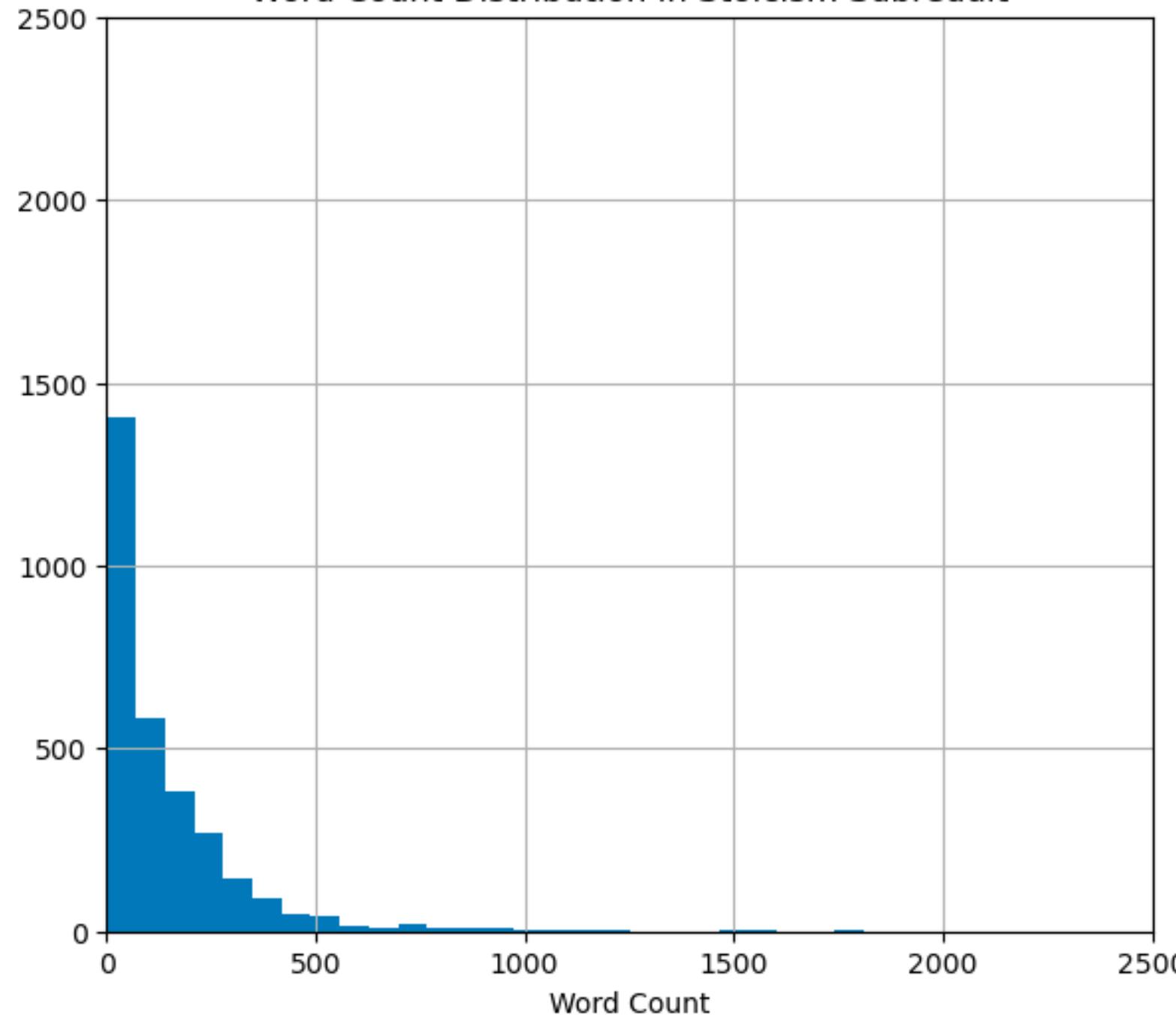
- Removed duplicates
- Combined title body text create 'combined_text' feature column
- Created target variable is_stoicism = 1
- Found word counts / class balance
 - Most frequent words/bigrams/trigrams
 - Removed punctuation, URLs & Emojis

Buddhism:
3279 submissions
327,513 words
100 words/post

Stoicism:
3057 submissions
436,787 words
143 words/post

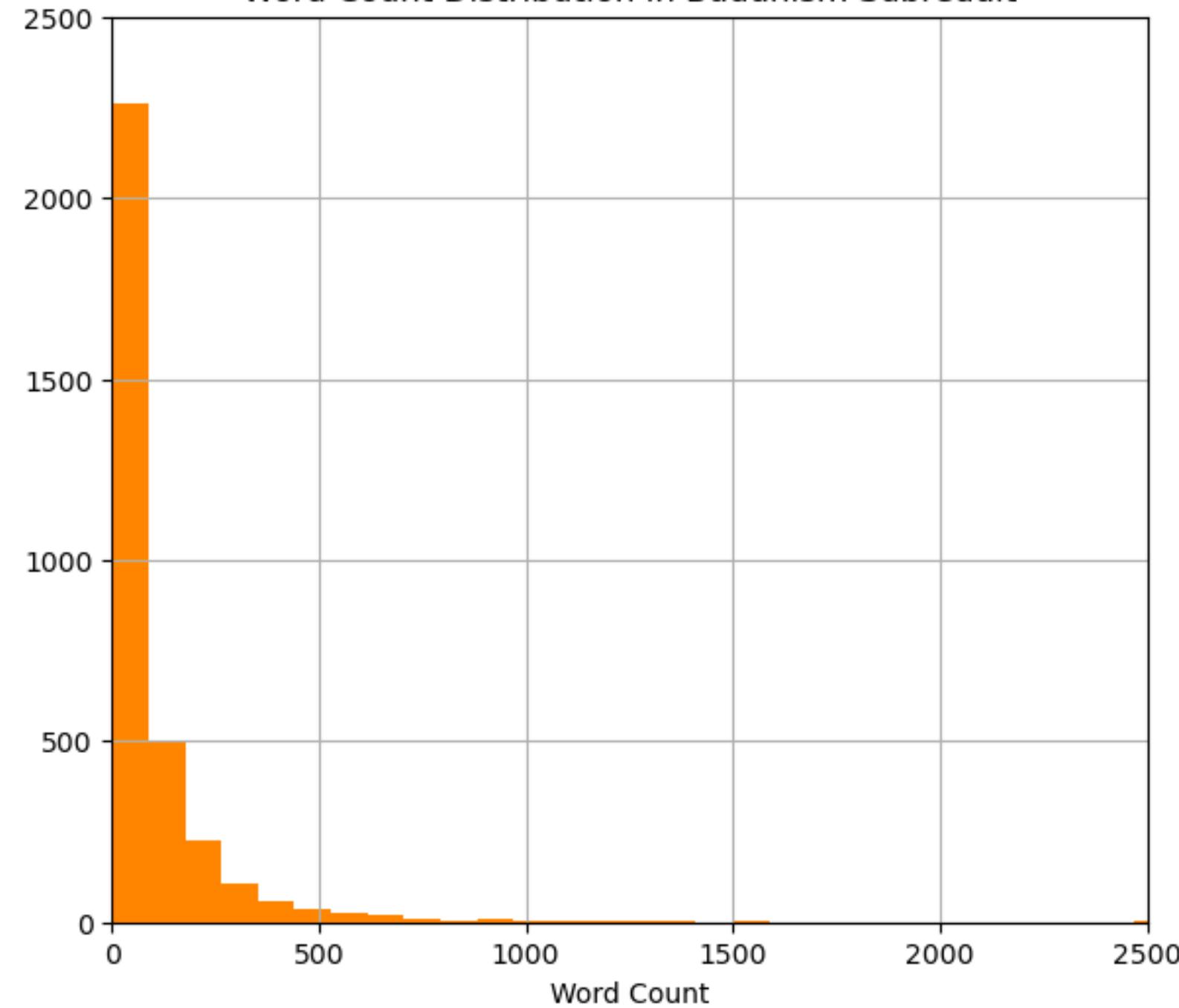
—

Word Count Distribution in Stoicism Subreddit



Median word count: 83

Word Count Distribution in Buddhism Subreddit



Median word count: 34

Random Forests



- Higher Variance
 - Train score: 0.99
 - Test score: 0.90
- Higher bias (cross validation) - 91%
- Less interpretability of coefficients

Production Model



TF-IDF

Ngram range: (1,1)

Max Features 4000

'English' stop words

Logistic Regression

$C = 0.97$

Performance Metrics

Logistic Regression with TF-IDF

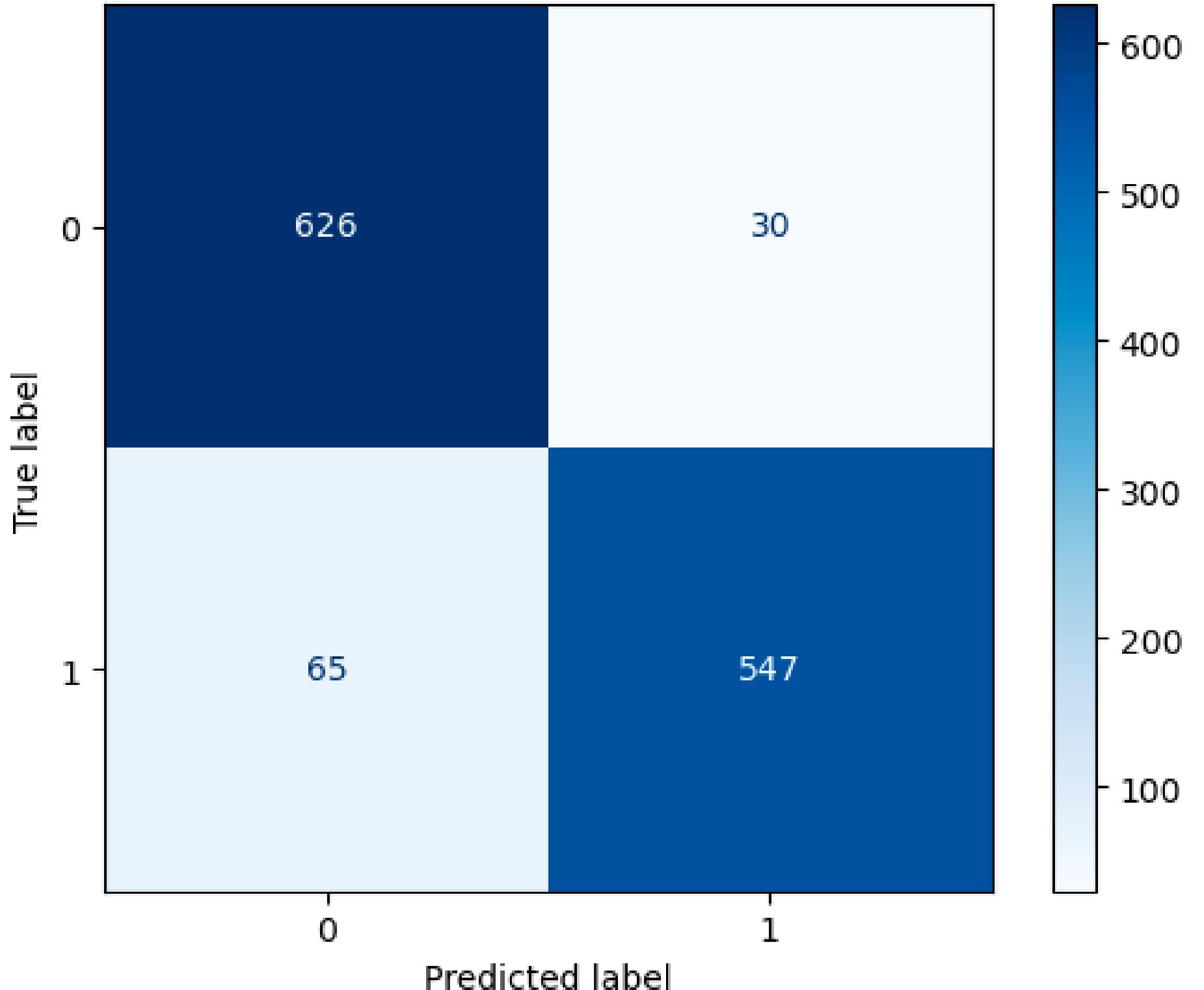
Cross Validation: 0.92

Train: 0.96, Test: 0.93

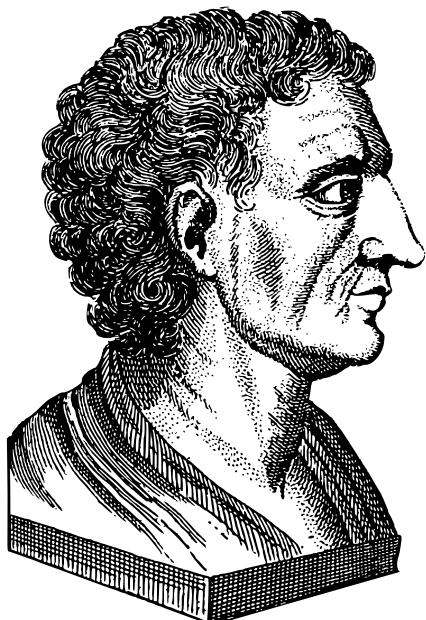
Balanced Accuracy: 0.92

Sensitivity (recall): 0.89

Specificity (precision): 0.95



Coefficient Analysis



Highest 15

Feature Name	IDF Score	Coefficient
stoicism	2.805865	9.206516
stoic	2.816728	8.543131
stoics	4.058628	4.129023
epictetus	4.278625	4.111882
marcus	3.810587	4.059558
aurelius	3.885452	3.489320
seneca	4.718714	3.349736
control	3.670113	3.184361
meditations	4.263041	2.722286
man	3.885452	2.454993
quote	4.507018	2.384863
philosophy	3.906881	2.059201
virtue	4.735108	1.665734
things	2.833865	1.647046
senecas	6.235062	1.512132

Lowest 15

Feature Name	IDF Score	Coefficient
buddhism	3.005869	-7.200463
buddhist	3.102794	-6.906719
buddha	3.344690	-5.991547
karma	4.702585	-3.396250
meditation	3.958745	-3.005583
buddhists	4.412905	-2.637770
temple	4.876938	-2.408037
dharma	4.710617	-2.382481
lama	5.174190	-2.144529
enlightenment	4.694617	-2.091396
monk	4.935779	-2.090478
practice	3.562191	-2.057617
buddhas	4.718714	-1.929221
rebirth	4.956188	-1.851860
compassion	4.794700	-1.826280

*toicism', 'stoic', 'stoics', 'rstoicismepictetus', 'seneca', 'marcus', 'aurelius', 'stupa', 'tantra',
'mori', 'memento', 'ryan', 'holiday', 'senecas', 'buddhism', 'buddha', 'buddhist', 'sutra',
'buddhists', 'buddhas', 'karma', 'lama', 'ajahn', 'rinpoche', 'bodhisattva', 'guanyin', 'meditation'
'rbuddhism', 'dhamma', 'altar', 'dalai', 'dharma', 'epictetus', 'hanh', 'mala', 'nhat', 'meditations',
'nirvana', 'sangha', 'sutra', 'zen', 'zeno', 'chrysippus', 'musonius', 'cleanthes', 'discourses',
'enchoridion', 'eudaimonia', 'logos', 'amor', 'fati', 'apatheia', 'prohairesis', 'cosmos', 'cynicism',
'dalai', 'thich', 'nhat', 'hanh', 'pali', 'canon', 'tripitaka', 'dhammapada', 'mahayana', 'theravada',
vajrayana', 'tibetan', 'c
'metta', 'bodhi', 'bodh
'mudra', 'tanha', 'vi
nandala', 'lamasery', 'l
na', 'zazen', 'samatha',
ata', 'koan', 'mantra',
da', 'sutra', 'tantra',
reen tara', 'white', tara*

Restricted Model

*'amitabha', 'maitreya', 'asanga', 'nagarjuna', 'vasubandhu', 'atisha', 'milarepa', 'tsongkhapa',
'shantideva', 'ashoka', 'vimalakirti', 'padmasambhavastoicism', 'rstoicism', 'zeno', 'chrysippus',
'musonius', 'cleanthes', 'discourses', 'enchoridion', 'eudaimonia', 'logos', 'amor', 'fati', 'apatheia',
'prohairesis', 'cosmos', 'cynicism', 'dalai', 'thich', 'nhat', 'hanh', 'pali', 'canon', 'tripitaka',
dhammapada', 'mahayana', 'theravada', 'vajrayana', 'tibetan', 'anatta', 'anicca', 'dukkha', 'samsara',
nirvana', 'vipassana', 'zazen', 'samatha', 'metta', 'bodhi', 'bodhisattva', 'dharma', 'skandha', 'prajna',
sunvata', 'tathagata', 'koan', 'mantra', 'mudra', 'tanha', 'vinaya', 'dana', 'sila', 'bhikkhu', 'bhikkhuni'*

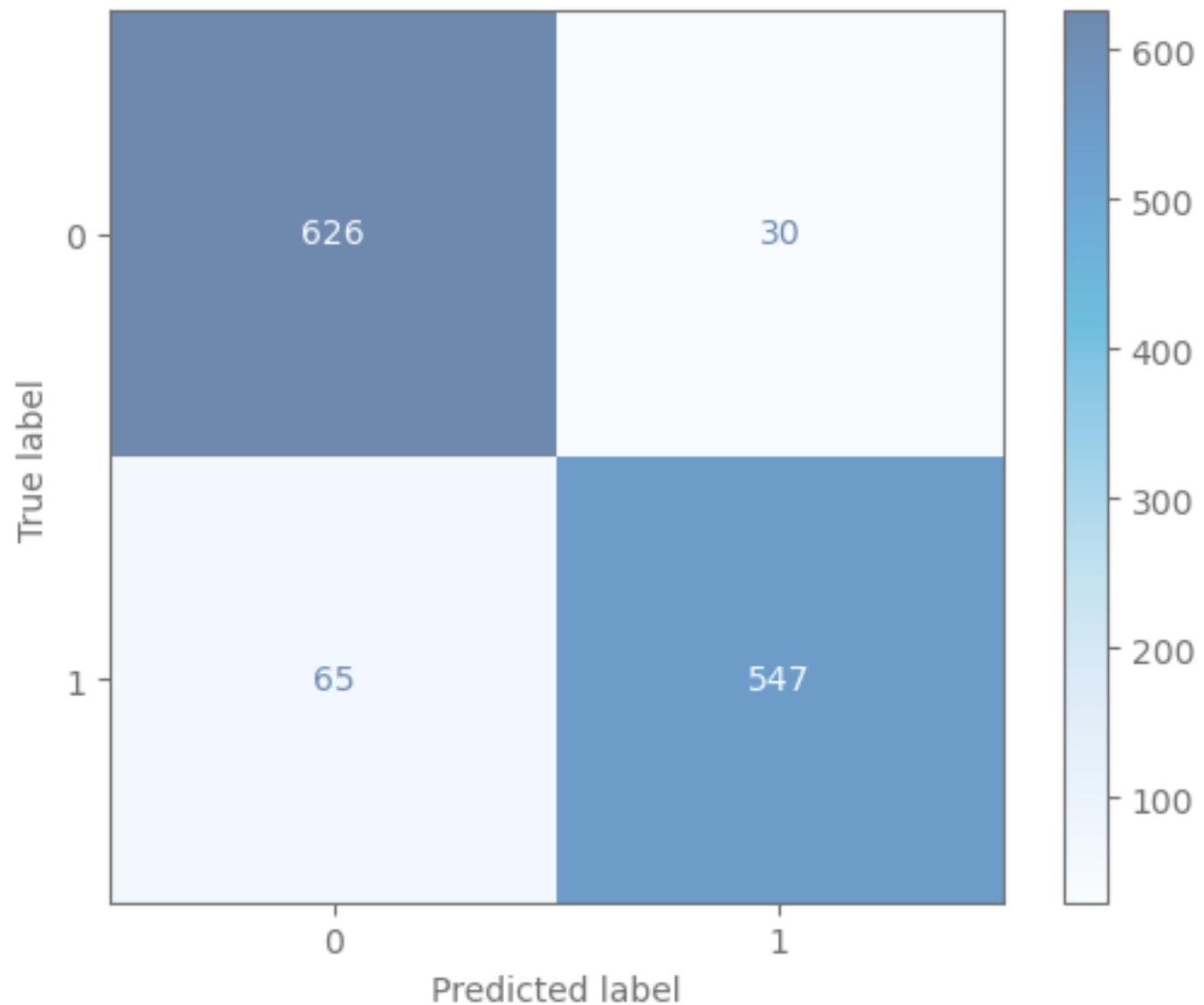
Production Model:

Cross Validation: 0.92

Balanced Accuracy: 0.92

Sensitivity: 0.89

Specificity: 0.95



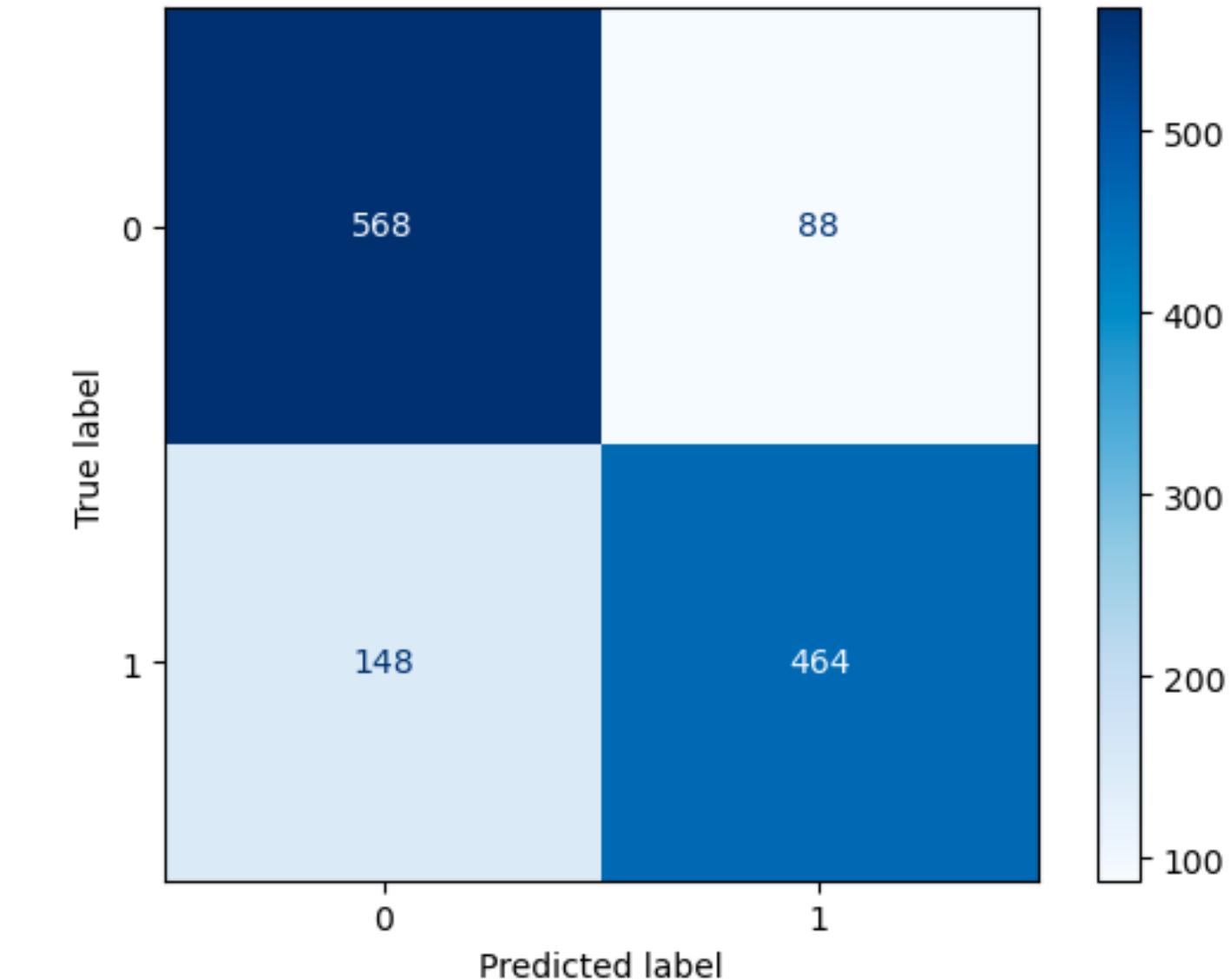
Restricted Model:

Cross Validation: 0.81

Balanced Accuracy: 0.81

Sensitivity: 0.76

Specificity: 0.86



r/Stoicism

Feature Name	IDF Score	Coefficent
control	3.670113	4.042794
philosophy	3.906881	3.872294
man	3.885452	3.074935
virtue	4.735108	2.892873
quote	4.507018	2.849824
quotes	5.254233	2.132224
emotions	4.395100	2.093007
live	3.664431	1.727812
letters	5.724236	1.708734
fate	5.523566	1.699172
things	2.833865	1.647737
power	4.500461	1.630817
life	2.572450	1.620589
virtuous	5.296792	1.599681
care	4.232581	1.594450
recommend	5.326206	1.557582
react	5.541915	1.537430
emotional	4.839551	1.534287
time	2.816728	1.533354
work	3.514742	1.521356

Verbs

control, live, react,
work, gonna,
respond, act, deal,
think, going, handle,
stop, advice

thank, practice,
suffering, teaching,
painting

Different Teachings

philosophy,
principles, virtue,
virtuous, fate,
power, care

rebirth,
enlightenment,
compassion,
teachings, beings,
tradition,
reincarnation

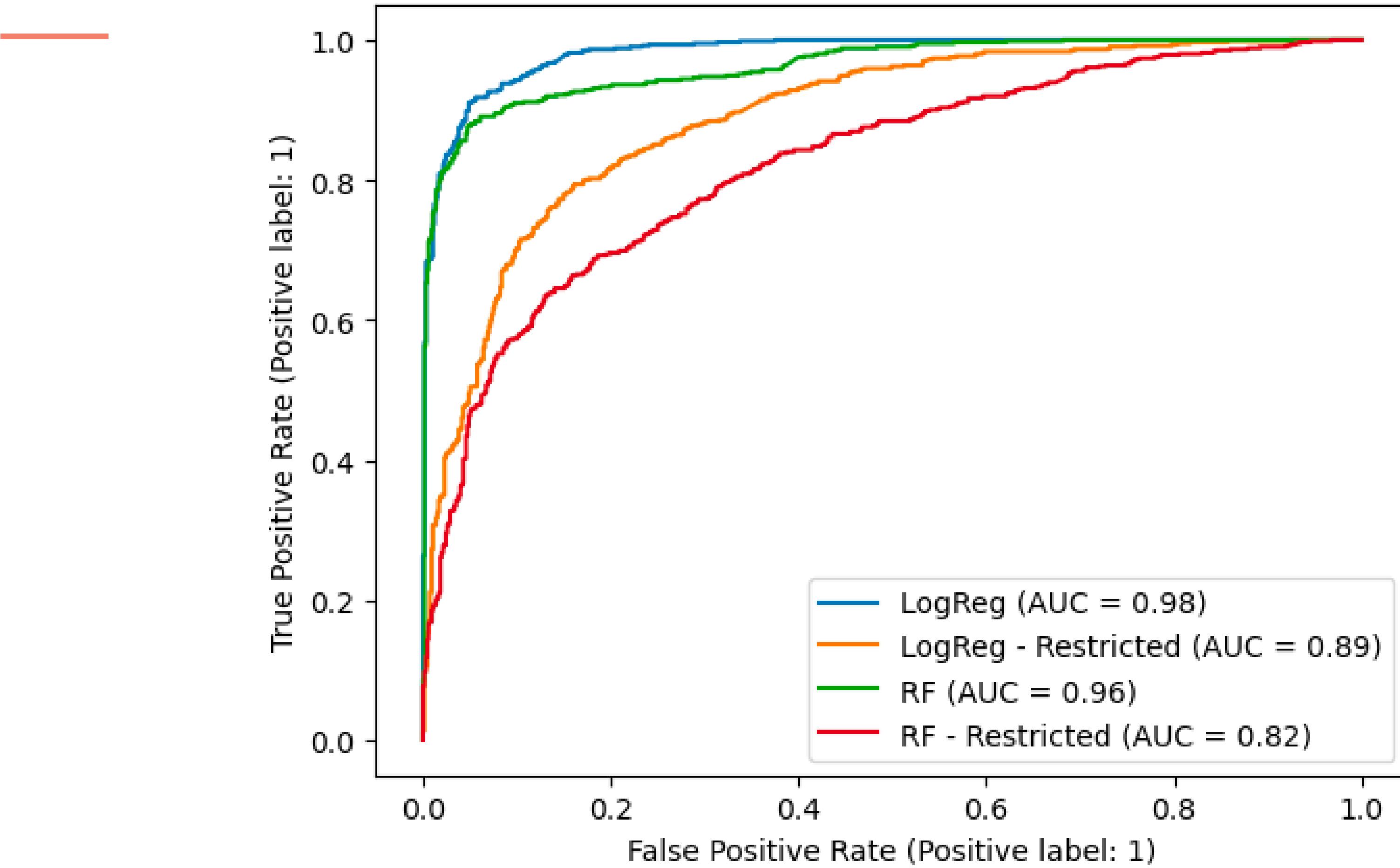
Mental States & Emotions

emotions,
emotionally,
emotion, react,
strength

suffering, peace,
compassion,
emptiness, mindful,
enlightenment,
enlightened, beautiful

r/Buddhism

Feature Name	IDF Score	Coefficent
temple	4.876938	-3.031032
rebirth	4.956188	-2.951588
teachings	4.193361	-2.894370
enlightenment	4.694617	-2.886637
practice	3.562191	-2.587546
compassion	4.794700	-2.458395
monastery	5.746709	-2.217291
statue	5.187093	-2.191435
path	4.160261	-1.989015
enlightened	5.436554	-1.924674
suffering	3.888992	-1.918014
beings	4.633059	-1.896112
teaching	4.858070	-1.807156
white	5.793229	-1.726656
shrine	6.395405	-1.579938
thank	3.800799	-1.550439
india	5.947380	-1.548300
reincarnation	5.724236	-1.537484
tradition	5.420025	-1.519947
pure	5.088248	-1.518677



Conclusion

SUCCESSFUL CLASSIFICATION MODEL MADE

92% accuracy on unseen data

RELIED QUITE HEAVILY ON TECHNICAL TERMS

Philosophers names,
technical terminology,
names of writings and
concepts.

DIFFERENCES FOUND

Stoics speak of '**virtue**' the
Buddhists '**enlightenment**'

'**care**' and '**compassion**'

REMOVING THESE TERMS

Decreased accuracy by 11% -
30% above the baseline.

REMARKABLY DISTINCT VOCABULARY

References

<https://philosophyasawayoflife.medium.com/stoicism-and-buddhism-a-comparison-58d3e86587b>

<https://thestoicsage.com/stoicism-and-buddhism/> <https://classicalwisdom.com/philosophy/stoicism-and-buddhism-two-sides-of-the-same-coin/>

<https://www.psychologytoday.com/intl/blog/buddhist-psychology-east-meets-west/202207/the-happy-overlap-between-stoicism-and-buddhism>

<https://www.quora.com/Why-isnt-Stoicism-considered-a-religion-since-Epictetus-mentions-God-so-much-in-his-Discourses>

<https://dailystoic.com/stoicism-buddhism/>

<https://rationallyspeaking.blogspot.com/2013/02/buddhism-epicureanism-and-stoicism.html>

<https://www.freecodecamp.org/news/how-to-write-a-regular-expression-for-a-url/>

<https://stackoverflow.com/questions/36182502/add-stemming-support-to-countvectorizer-sklearn>

<https://stackoverflow.com/questions/161738/what-is-the-best-regular-expression-to-check-if-a-string-is-a-valid-url?page=2&tab=votes>

<https://gist.github.com/slowkow/7a7f61f495e3dbb7e3d767f97bd7304b>

<https://www.includehelp.com/python/fast-punctuation-removal-with-pandas.aspx> <https://stackoverflow.com/questions/5511708/adding-words-to-nltk-stoplist> <https://towardsdatascience.com/how-to-visualize-a-decision-tree-from-a-random-forest-in-python-using-scikit-learn-38ad2d75f21c>

<https://towardsdatascience.com/decision-trees-explained-entropy-information-gain-gini-index-ccp-pruning-4d78070db36c>



Q & A