A Study of Ignorance

By Anand Ramakrishnan

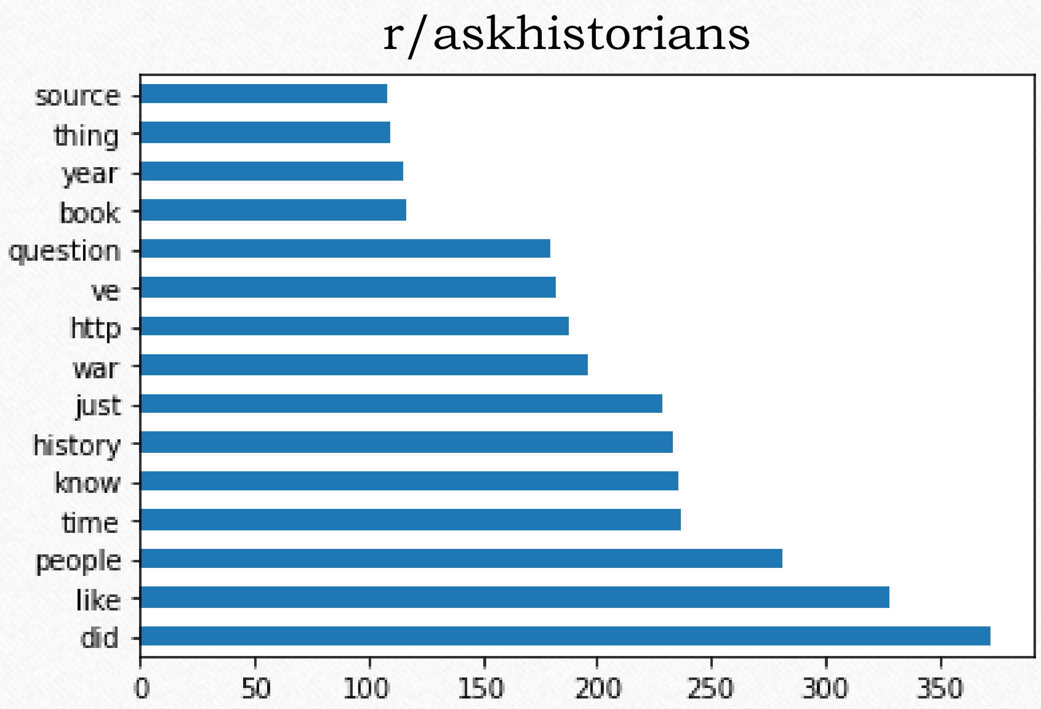
What are r/badhistory and r/AskHistorians?

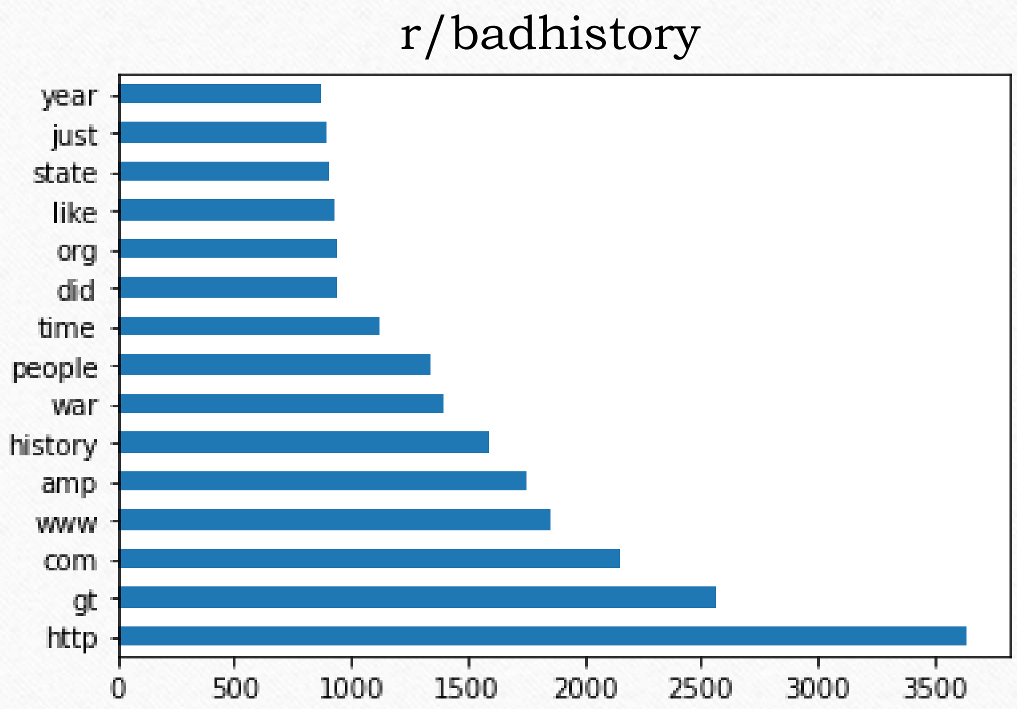
- r/badhistory is a subreddit dedicated to rebuttals of those who present false or misleading historical information.
- r/AskHistorians is a subreddit where users ask academic-level questions about history.

Problem Statement

- Since the posts in r/badhistory are about those who are **willfully wrong**, while the posts in r/AskHistorians are made by those who are merely **uninformed**, the ignorance highlighted in r/badhistory is more dangerous.
- We should try to classify posts based on which subreddit they come from, as the results will give a clue as to which historical terms are more associated with wrongness.
- The point of this exercise is to figure out which model is best for the classification.

Top Words


r/askhistorians
source
thing
year
book
question
ve
http
war
just
history
know
time
people
like
did
0
50
100
150
200
250
300
350

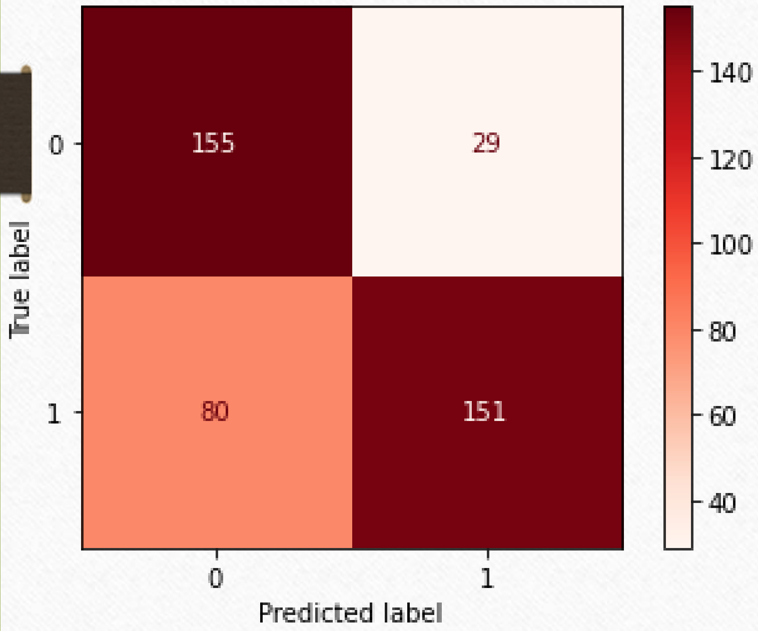

r/badhistory
year
just
state
like
org
did
time
people
war
history
amp
www
com
gt
http
0
500
1000
1500
2000
2500
3000
3500

Three Good Models

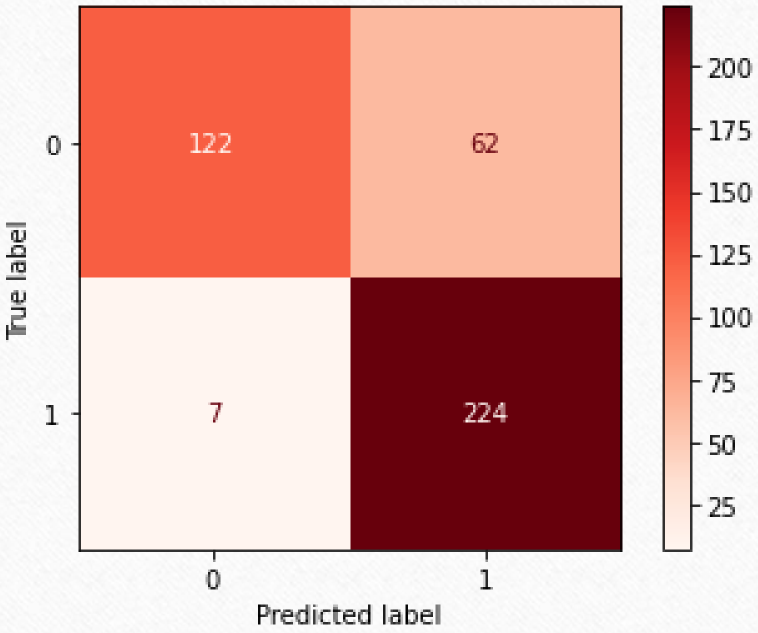
Model	Training Set Score	Test Set Score	Sensitivity	Specificity	Precision
Multinomial Naïve Bayes	0.905	0.737	0.653	0.842	0.838
Logistic Regression	0.986	0.833	0.969	0.663	0.783
Random Forest	0.997	0.824	0.826	0.82	0.852

Confusion Matrices

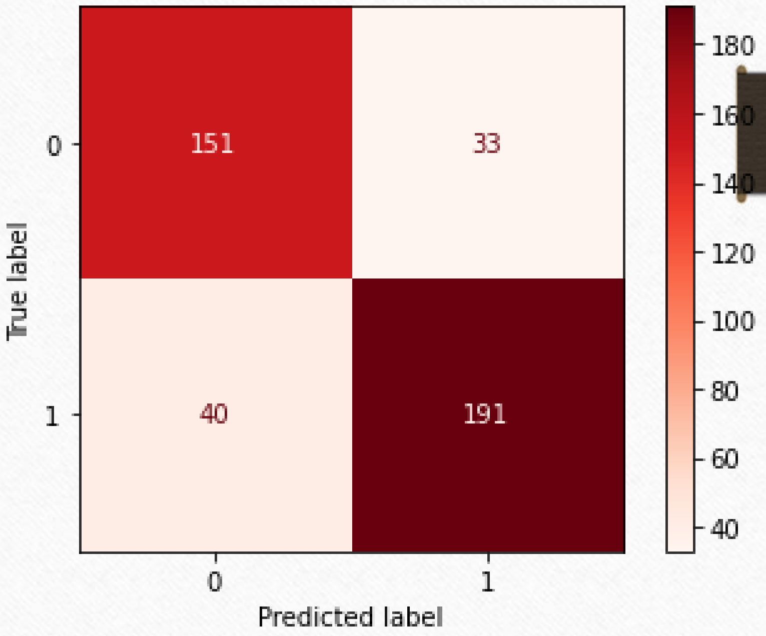
Naïve Bayes



Logistic Regression



Random Forest



Conclusion & Further Study

- Since the point of this problem is to separate those who are **wrong** about history, it is more important that I correctly identify those in the former category. Thus, it would be better to have higher specificity than higher sensitivity.
- Given both the high specificity and accuracy, I would argue to use the **random forest classifier** model to determine whether a post is in r/AskHistorians or r/badhistory based on its contents.
- As r/badhistory contains so many posts with links, future research would need to show if an algorithm can clearly tell non-link posts in r/badhistory from those in r/AskHistorians.