Political Tweet Classifier

By Arsh Patel

My Idea

To see if you can create a model that can predict whether political tweets belong to a certain affiliation (left-wing OR right-wing)

Only text was used as a feature of the data

My Data

Used SNScraper to get tweets

1/20/2017 onwards (Trump Era)

Scraped 146K Dem tweets and 138K GOP tweets , 285K altogether

Tweets from Congressmen and the last two Presidents

My Data

Used both personal and professional twitter accounts of these politicians

Tried to balance between D and R

Democrats	Republicans	
Joe Biden	Donald Trump	Jim Jordan
Kamala Harris	Donald Trump Jr.	Lauren Boebert
Nancy Pelosi	Mike Pence	@ GOP
Chuck Schumer	Ted Cruz	
Bernie Sanders	Marco Rubio	
Elizabeth Warren	Mitch McConnell	
Alexandria Ocasio-Cortez	Kevin McCarthy	
Maxine Waters	Paul Gosar	
Cory Booker	Marjorie Taylor Greene	
Pramila Jayapal	Rand Paul	
Amy Klobuchar	Tom Cotton	
@TheDemocrats	Rick Scott	

Text Preprocessing

Standard NLP preprocessing for tweets

Only key difference = custom stopwords

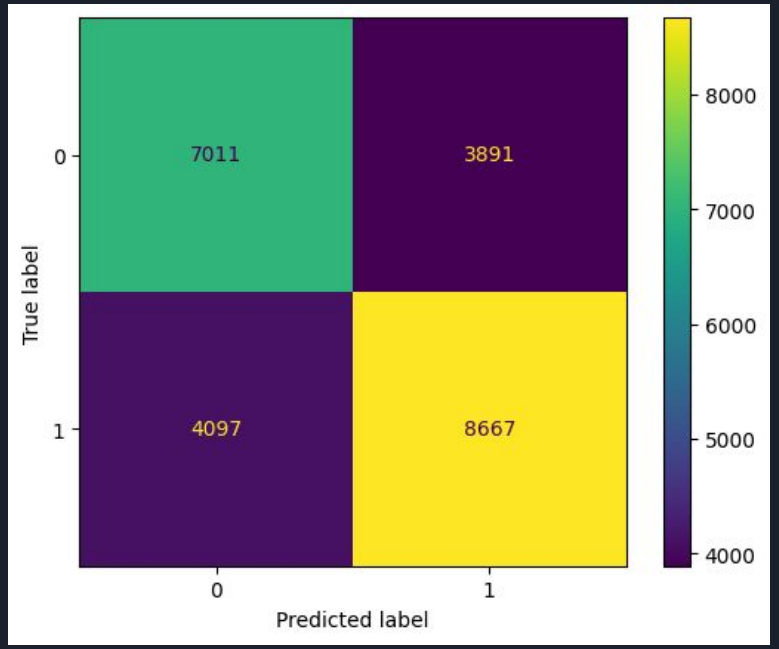
Improved model score by 1.5 - 2.5%

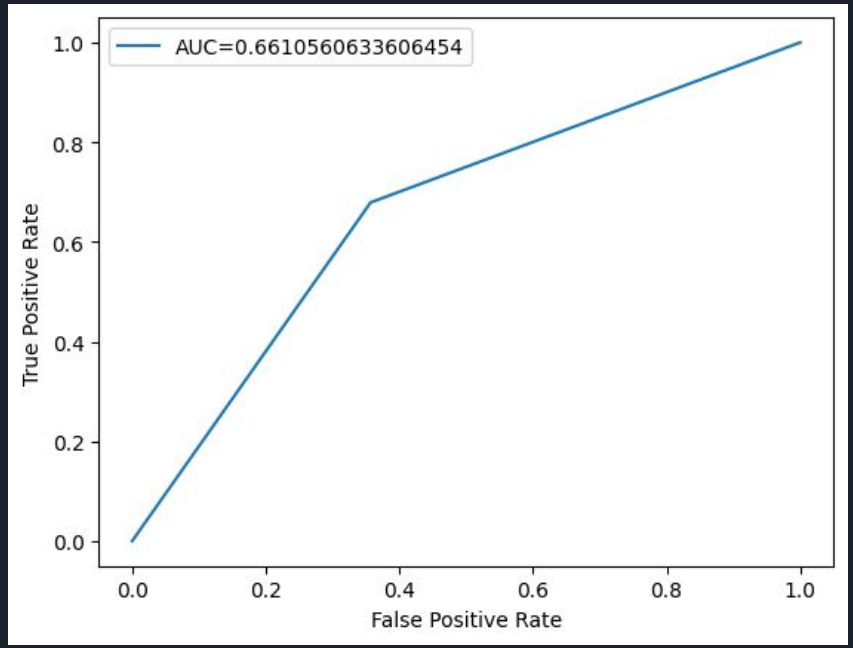
Only used tweets with at least 8 words

Final data had 237K tweets

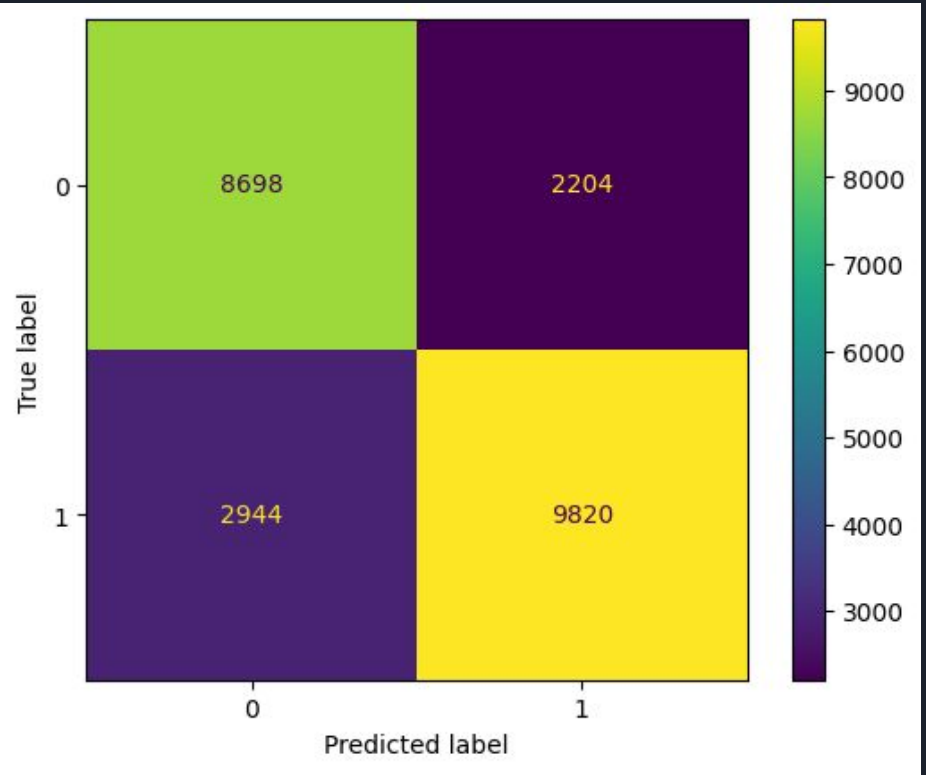
Baseline Model

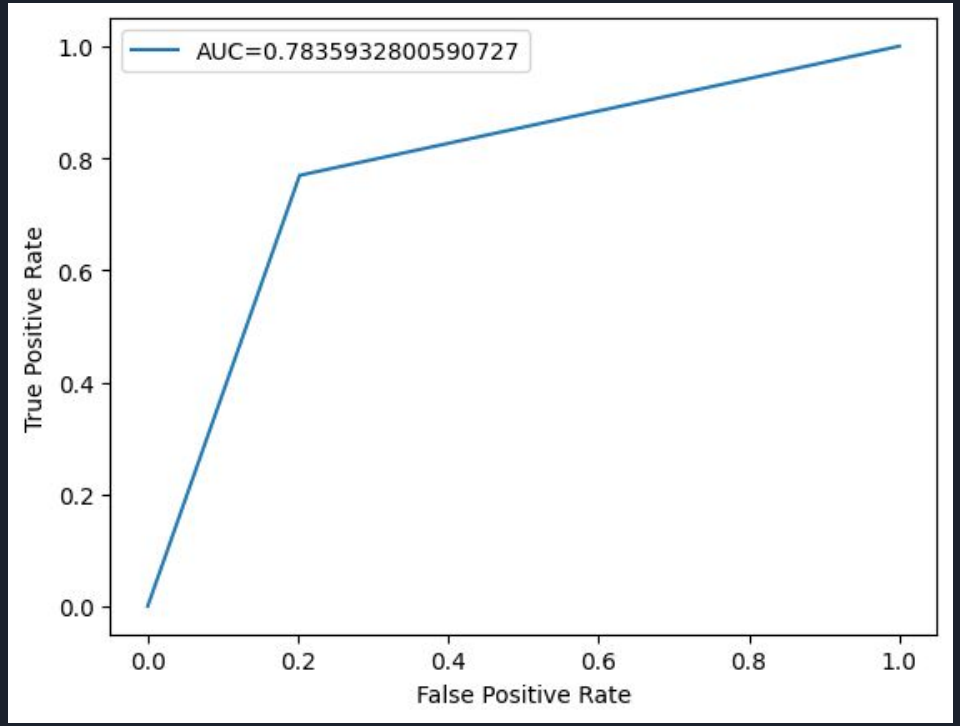
Multinomial Naive Bayes with TfidfVectorizer





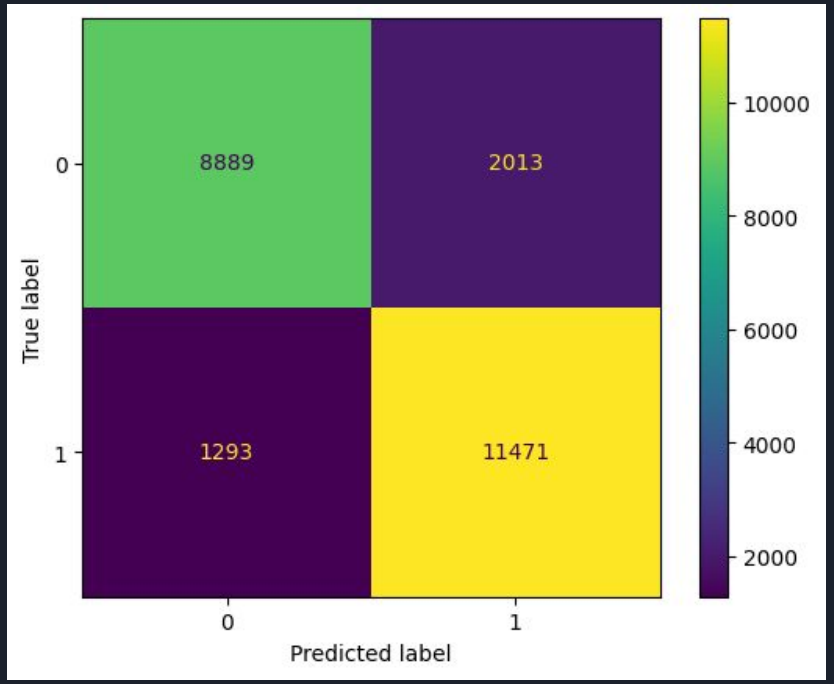
Model Tuning the Naive Bayes Model

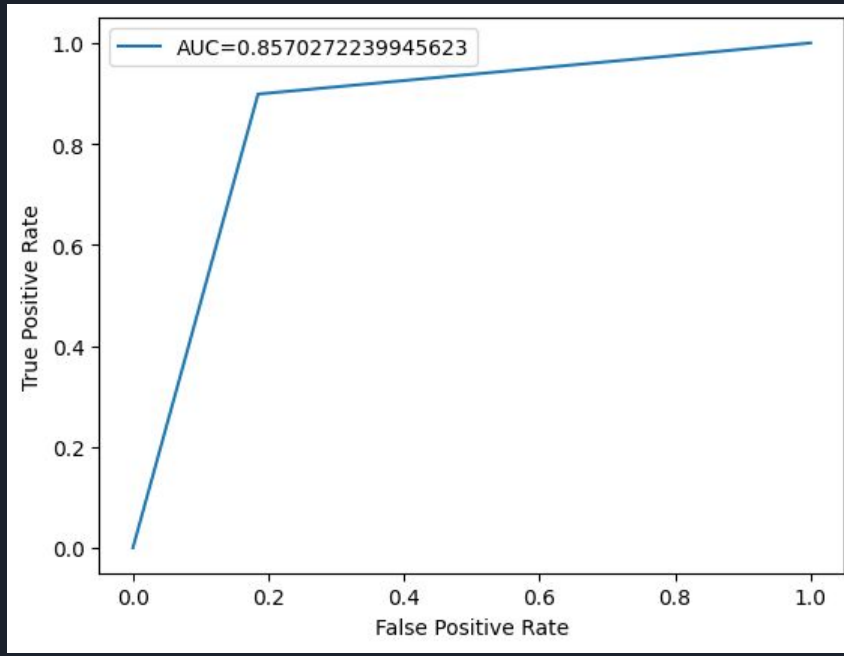




Final Model...so far

Hypertuned Random Forest





Conclusion

Model has been successfully created that can classify tweets from politicians as left-wing or right-wing based on text

Final model: Hypertuned Random Forest with ~85% accuracy

Next Steps

Do Sentiment Analysis on the tweets, see how they differ based on ideology

Expand project to make this a categorical classifier

Use other models, like neural networks, to see if they can replicate the success of the random forest model