

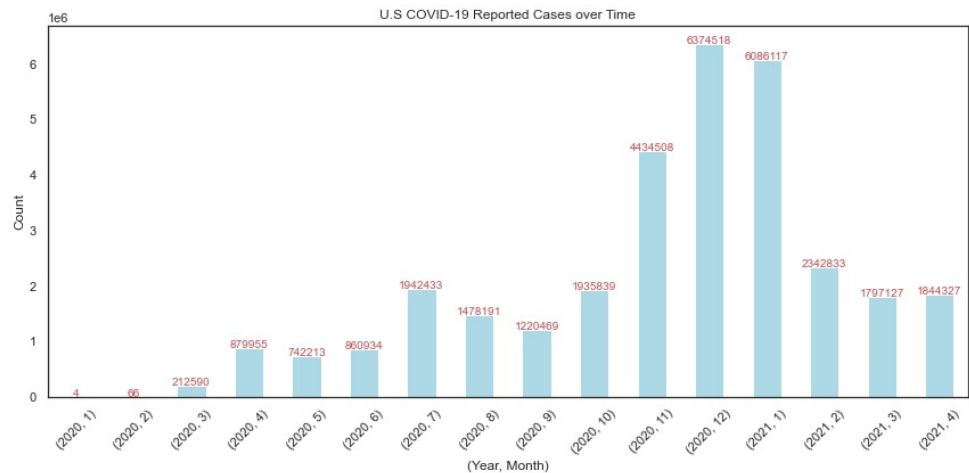
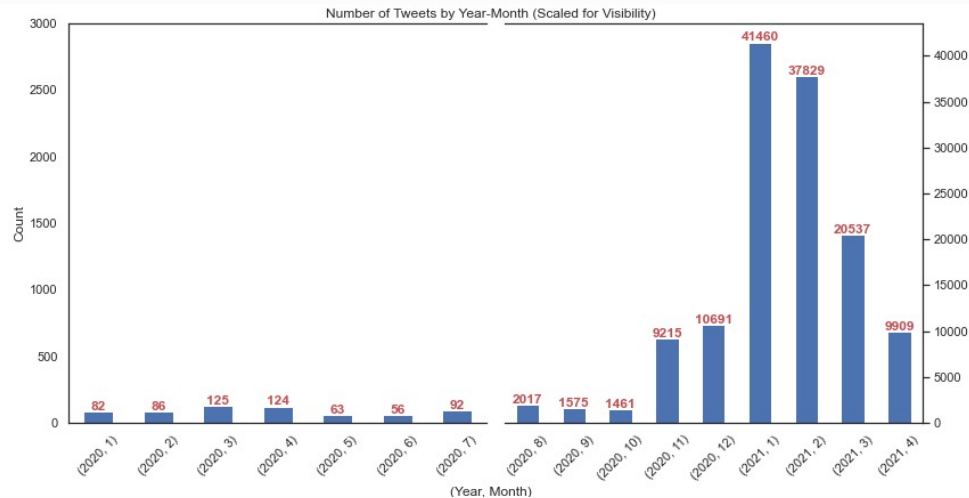
COVID-19 Vaccine Twitter Sentiment Analysis



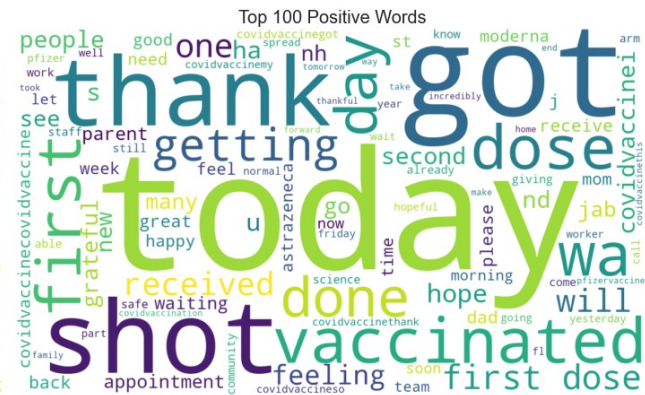
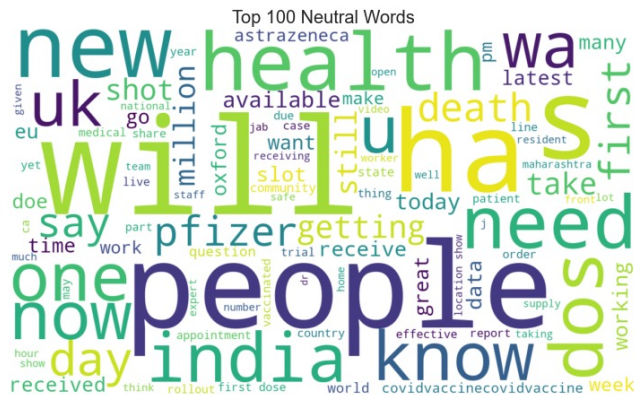
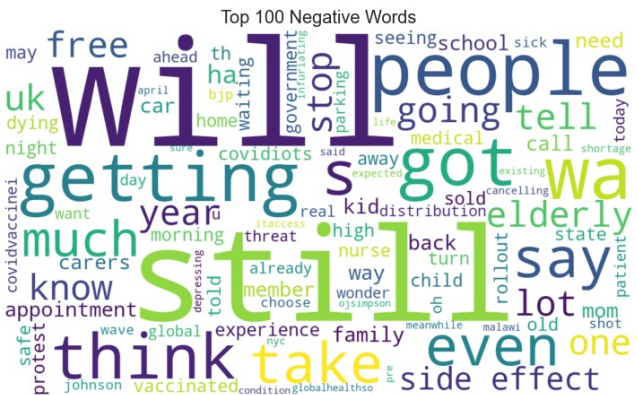
Billie Kim

Dec. 2021

Date Trends



Copyright 2019 Pearson Education, Inc. or its affiliate(s). All rights reserved.



Annotations

Negative



"So, in our limited experience, this vaccine is quite clearly more dangerous than covid19 just saying folks be warned be prepared."

Neutral



"I don't know what to believe or do anymore about this covidvaccine ppl are saying take the shot it could help not get a sick with deltavarient then ppl saying the vaccine taker will be targeted with delta I am confused."

Positive



"It is so important that we all receive the covid vaccine if we are to beat or even keep this disease at bay we are so lucky to have our amazing help them to help you get your jab stay safe be covidsafe."

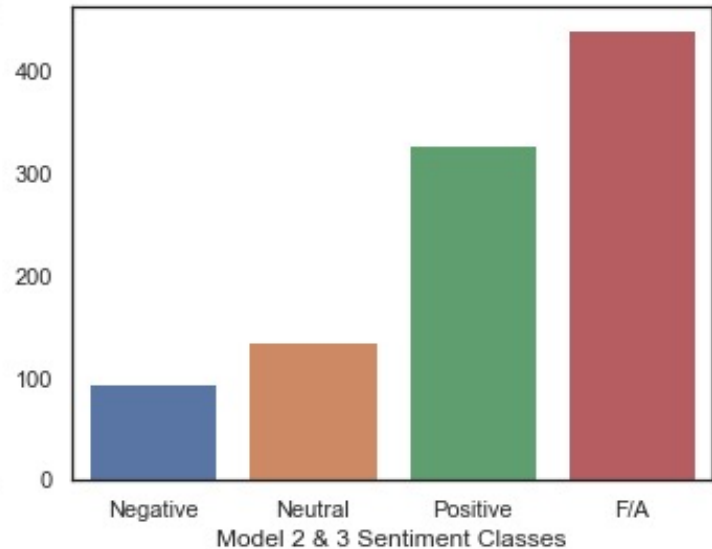
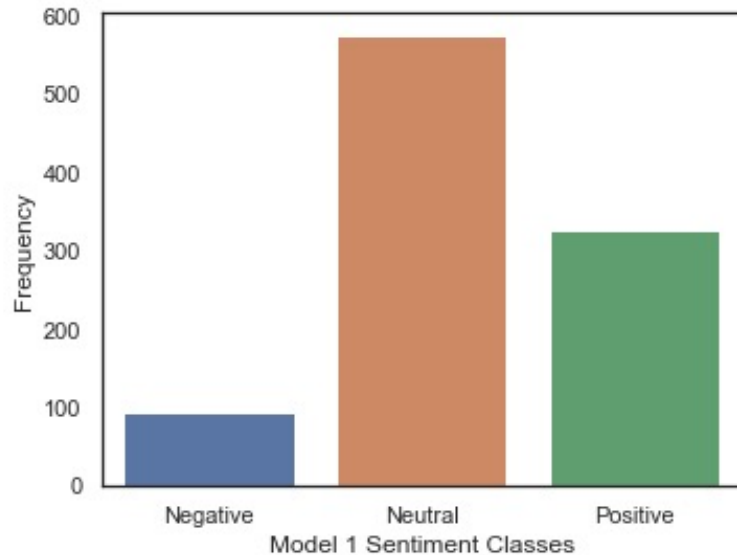
Influential Terms

y=Negative top features		y=Neutral top features		y=Positive top features	
Weight?	Feature	Weight?	Feature	Weight?	Feature
+1.015	still	+1.006	<BIAS>	+2.350	my
+0.904	not	+0.708	he	+1.757	covidvaccine
+0.899	if	+0.681	by	+1.568	got
+0.820	cannot	+0.658	about	+1.406	dose
+0.731	they	+0.657	show	+1.323	vaccinated
+0.619	only	+0.653	health	+1.215	done
+0.582	because	+0.608	from	+1.123	today
+0.575	sold	+0.607	india	+1.064	first
+0.569	america	+0.529	what	+0.980	hope
+0.561	are	+0.514	open	+0.904	thank
+0.546	elderly	+0.497	million	+0.900	grateful
+0.521	no	+0.484	your	+0.899	so
+0.519	think	+0.463	dr	+0.831	shot
+0.504	maybe	+0.460	on	+0.801	nh
+0.500	wonder	+0.449	are	+0.790	feeling
+0.497	yes	+0.449	national	+0.762	she
+0.489	free	+0.449	oxford	+0.739	waiting
+0.489	dying	+0.437	available	+0.730	thanks
+0.485	covidiots	+0.417	new	+0.727	see
+0.477	most	+0.414	death	+0.717	just
... 467 more positive 1553 more positive 744 more positive ...	
... 2078 more negative 992 more negative 1801 more negative ...	
-0.387	your	-0.862	just	-0.451	show
-0.417	coronavirus	-0.871	am	-0.456	effect
-0.421	my	-0.873	waiting	-0.461	on
-0.427	today	-0.910	still	-0.478	from
-0.449	covid	-1.013	done	-0.488	how
-0.496	at	-1.026	dose	-0.527	health
-0.497	first	-1.091	vaccinated	-0.529	not
-0.516	for	-1.541	got	-0.549	say
-0.600	vaccine	-1.605	covidvaccine	-0.582	about
-0.869	<BIAS>	-1.929	my	-1.010	are

Preprocessing Steps

1. HTML decoding — HTML encoding will be decoded into standard English characters
2. Removing twitter handles (@mentions) and URL links
3. Expand contractions
4. Lower-casing
5. Removing numbers and special characters
6. Tokenization
7. Lemmatization

Sentiment Classes



Modeling Approaches

1

Model 1: Restricted/Nested Model

- Sentiment Classes: Negative, Neutral, Positive
- Achieved 81.3% Test Accuracy

2

Model 2: Full Model

- Sentiment Classes: Negative, Neutral, Positive, F/A
- Achieved 70.7% Test Accuracy

3

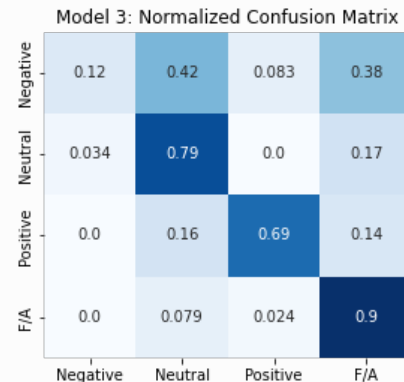
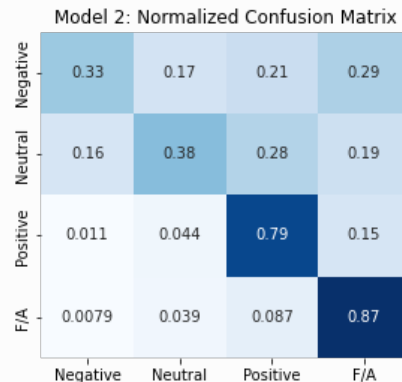
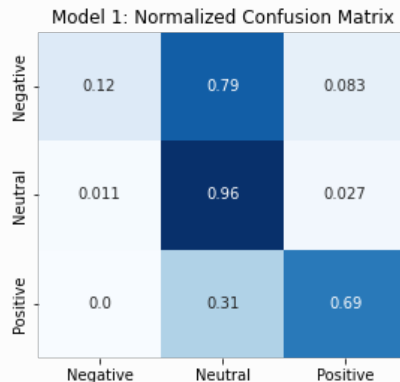
Model 3: 2-Step Model

- Sentiment Classes: Negative, Neutral, Positive, F/A
- Step 1: Subjectivity Classifier
- Step 2: Sentiment Classifier
- Achieved 75.3% Test Accuracy

Model 3: Steps 1 & 2

	Test Acc.	F1 (macro)	F1 (weighted)	Error-Rate
(Step 1) Subjectivity Classifier Algorithm				
Logistic Regression	0.817	0.808	0.814	0.183
Ridge Classifier	0.820	0.814	0.819	0.180
Voting Ensemble	0.843	0.842	0.844	0.157
	Test Acc.	F1 (macro)	F1 (weighted)	Error-Rate
(Step 2) Full System Algorithm Combination				
Logistic Regression LR	0.693	0.588	0.695	0.307
Ridge Classifier LR	0.700	0.591	0.699	0.300
Voting Ensemble LR	0.753	0.624	0.738	0.247

Conclusions & Findings



Model	Algorithm / Method	Test Accuracy	Null Accuracy	Δ Acc.	Error-Rate
Model 1: Reduced Sentiment Model	Logistic Regression	0.813	0.579	+23.43%	0.187
Model 2: Full Sentiment Model	Logistic Regression	0.707	0.427	+27.97%	0.293
Model 3: 2 Step Model	Logistic Regression LR	0.693	0.427	+26.63%	0.307
	Ridge Classifier LR	0.700	0.427	+27.3%	0.300
	Ensemble Voting LR	0.753	0.427	+32.63%	0.247