Twitter US Airline Sentiment Analyzer using Machine Learning

Loïc XU

Table of contents

1 - Introduction

Perform sentiment analysis on tweets related to flights

A door of Alaska Boeing 737 plane was ripped-off in mid-air

Airline companies may use social media to gauge the satisfaction of their customers

2 - Data

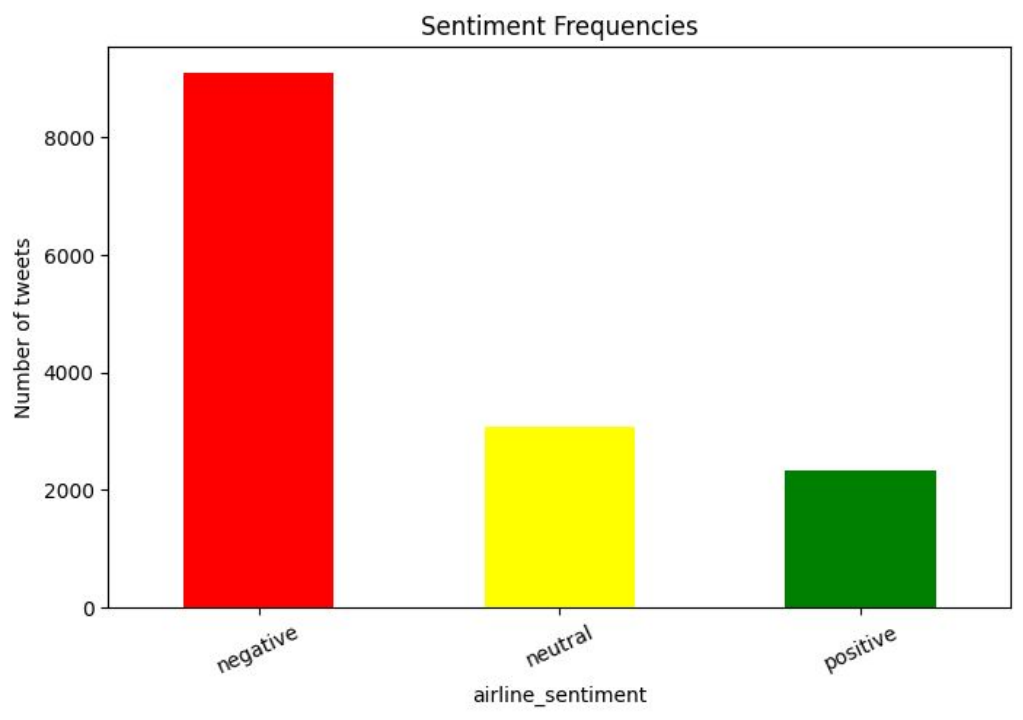
Dataset from Kaggle (2015)

Human labelled

14640 tweets with 15 columns

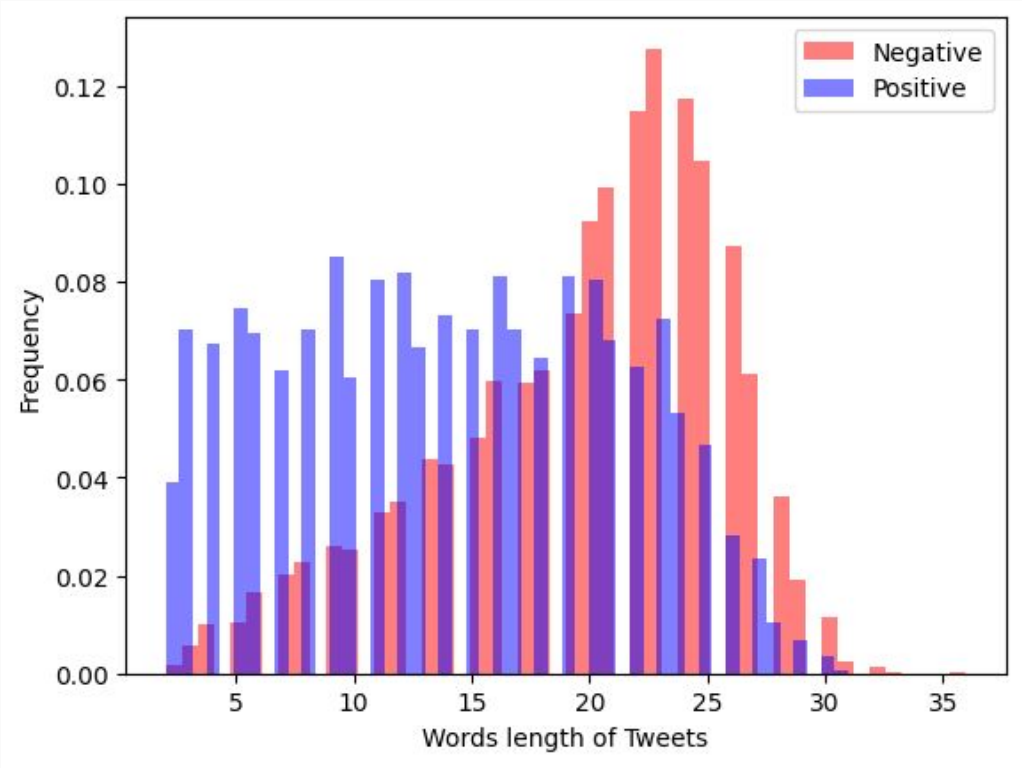
Negative or positive or neutral

2 - Data



Negative	Positive	Neutral
9082	3069	2334

2 - Data

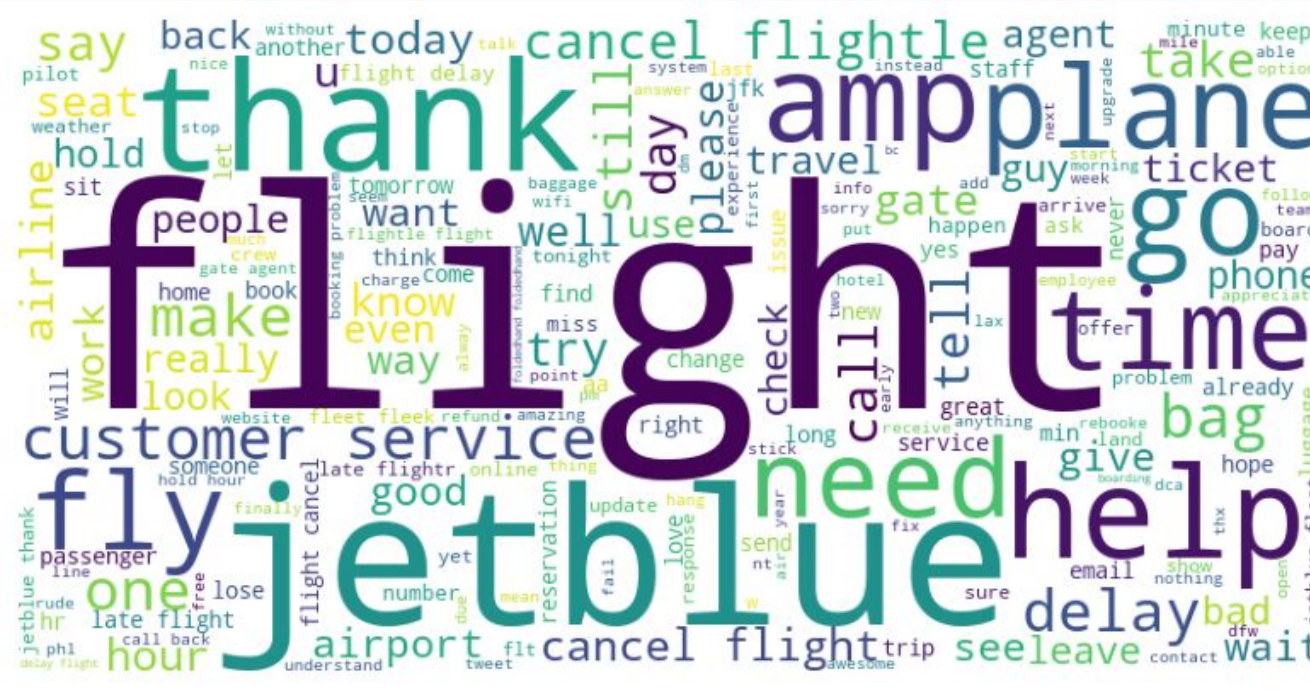

Negative
Positive
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Frequency
5
10
15
20
25
30
35
Words length of Tweets

3 - Pre-processing

1. remove duplication
2. convert emojis to text with demojize (emoji library)
3. tokenize the tweets (NLTK library)
4. remove stopwords and symbols
5. remove every company's name
6. lemmatize tokens (SpaCy library)
7. transform sentiment into numerical values

Tweet	Pre-processed tweet
I ❤️ flying @VirginAmerica. ☺👍	[redheart, fly, smilingface, thumbsup]

3 - Pre-processing



Wordcloud after pre-processing

3 - Pre-processing

Feature engineering :

1. TF-IDF
2. TF
3. Tweet’s length
4. N-gram
5. VADER (Valence Aware Dictionary and sEntiment Reasoner) compound score

4 - Models

Models	Baseline
Logistic Regression (OvR)	Naive Bayes
Random Forest	VADER
BERT	

5 - Results

Table 2: Classification Report for the baseline models

Model	Class	Precision	Recall	F1-Score
Random Guess	0	0.65	0.34	0.45
	1	0.14	0.32	0.19
	2	0.21	0.34	0.26
	Accuracy on test set			**0.3379**
	Accuracy on train set			0.3416
VADER	0	0.90	0.43	0.58
	1	0.26	0.92	0.41
	2	0.38	0.33	0.35
	Accuracy on test set			**0.4911**
	Accuracy on train set			0.4811
Always predicting negative	0	0.65	1.00	0.78
	1	0	0	0
	2	0	0	0
	Accuracy on test set			**0.6454**
	Accuracy on train set			0.6223
Naive Bayes TF-IDF	0	0.72	0.99	0.83
	1	0.85	0.32	0.46
	2	0.73	0.23	0.35
	Accuracy on test set			**0.7314**
	Accuracy on train set			0.7714

Table 5: Classification Report for the best Random Forest models

Model	Class	Precision	Recall	F1-Score
RF TF-IDF 2-gram	0	0.78	0.94	0.85
	1	0.79	0.44	0.56
	2	0.63	0.43	0.51
	Accuracy on test set			**0.7604**
	Accuracy on train set			0.9985
RF TF-IDF	0	0.79	0.92	0.85
	1	0.72	0.54	0.62
	2	0.62	0.43	0.51
	Accuracy on test set			**0.7621**
	Accuracy on train set			0.9861
RF TF 2-gram + TF-IDF 4-gram + TF-IDF + tweet length	0	0.78	0.95	0.86
	1	0.81	0.45	0.58
	2	0.66	0.43	0.52
	Accuracy on test set			**0.7694**
	Accuracy on train set			0.9989
RF TF 2-gram + TF-IDF 4-gram + TF-IDF + VADER score	0	0.77	0.94	0.85
	1	0.78	0.46	0.57
	2	0.64	0.41	0.50
	Accuracy on test set			**0.7680**
	Accuracy on train set			0.9989

5 - Results

Table 3: Classification Report for the best Logistic Regression models

Model	Class	Precision	Recall	F1-Score
LR TF-IDF	0	0.83	0.91	0.87
	1	0.74	0.63	0.68
	2	0.62	0.51	0.56
	Accuracy on test set			**0.7849**
	Accuracy on train set			0.9005
LR TF-IDF	0	0.84	0.90	0.87
+ tweet length	1	0.75	0.64	0.69
	2	0.62	0.53	0.57
	Accuracy on test set			**0.7880**
	Accuracy on train set			0.8992
LR TF 2-gram	0	0.86	0.90	0.88
	1	0.75	0.65	0.70
	2	0.64	0.61	0.62
	Accuracy on test set			**0.8046**
	Accuracy on train set			0.9981
LR TF 2-gram	0	0.87	0.91	0.89
+ TF-IDF 4-gram + TF-IDF	1	0.76	0.66	0.71
	2	0.65	0.62	0.63
	Accuracy on test set			**0.8101**
	Accuracy on train set			0.9984
LR TF 2-gram	0	0.87	0.92	0.89
+ TF-IDF 4-gram + TF-IDF	1	0.76	0.69	0.72
+ VADER score	2	0.68	0.61	0.64
	Accuracy on test set			**0.8194**
	Accuracy on train set			0.9909

Table 4: Classification Report for the BERT models

Model	Class	Precision	Recall	F1-Score
DistilBERT fine-tuned on SST2	0	0.77	0.90	0.83
	1	0.51	0.86	0.64
	2	0.00	0.00	0.00
	Accuracy on test set			**0.7079**
	Accuracy on train set			X
DistilBERT	0	0.73	0.02	0.04
	1	0.12	0.55	0.20
	2	0.22	0.34	0.27
	Accuracy on test set			**0.1615**
	Accuracy on train set			X
DistilBERT fine-tuned on the dataset	0	0.89	0.91	0.90
	1	0.77	0.81	0.79
	2	0.70	0.63	0.66
	Accuracy on test set			**0.8394**
	Accuracy on train set			*0.4591*
BERT fine-tuned on the dataset	0	0.91	0.92	0.91
	1	0.81	0.75	0.78
	2	0.67	0.69	0.68
	Accuracy on test set			**0.8453**
	Accuracy on train set			*0.4598*
RoBERTa fine-tuned on the dataset	0	0.93	0.89	0.91
	1	0.71	0.88	0.79
	2	0.71	0.68	0.70
	Accuracy on test set			**0.8463**
	Accuracy on train set			*0.4535*

5 - Results

thank	8.678980
hour	7.314860
delay	6.402677
bad	5.441299
lose	5.267737
not	5.179501
hrs	5.175680
amazing	5.048006
nothing	4.926525
awesome	4.862812
great	4.835204
cancel	4.740522
lack	4.659590
ridiculous	4.642117
rude	4.594183
luggage	4.518131
suck	4.449456
excellent	4.332189
late	4.312874
error	4.294429

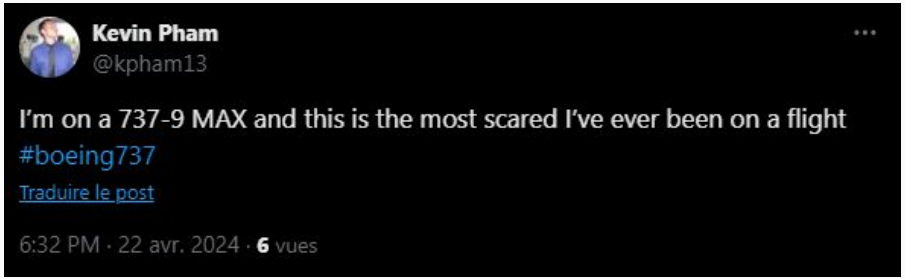
negative

great	8.324125
thank	7.923535
awesome	7.320832
amazing	6.675515
excellent	6.511010
love	6.508600
kudo	5.499933
thx	5.437003
rock	5.429575
wonderful	5.141357
hour	5.018146
appreciate	4.910096
good	4.835569
foldedhand	4.495737
excited	4.400565
thnx	4.375687
thankful	4.186679
need	4.161480
thumbsup	4.049082
delay	4.018652

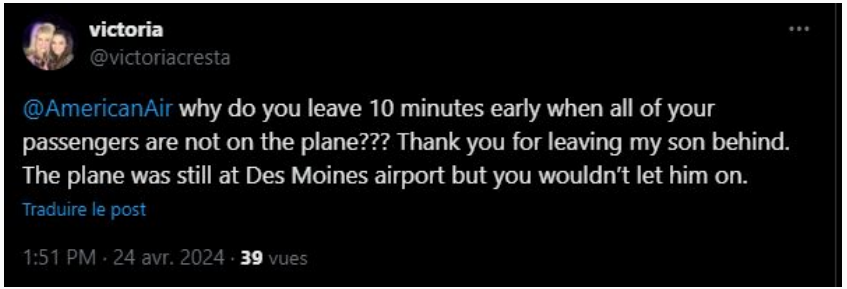
positive

hour	5.875150
lose	5.267706
delay	5.051412
experience	4.707122
fix	4.618432
helpful	4.572547
customer	4.570996
great	4.553321
bad	4.496677
agent	4.352999
cancel	4.313202
wait	4.251182
not	4.058446
attendant	4.045809
suck	4.006297
awesome	3.888679
unacceptable	3.820775
ridiculous	3.817301
frustrating	3.798576
error	3.794317

neutral

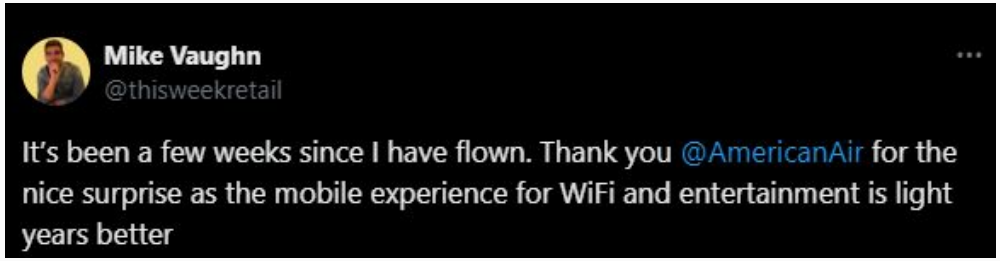


Label : 0



Label : 1

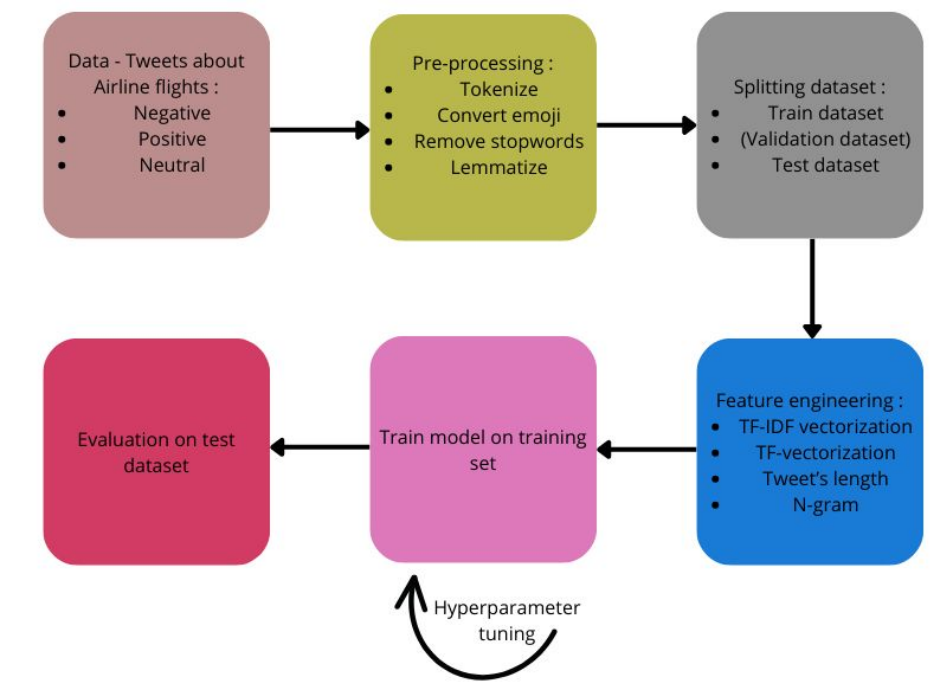
Label : 0



Label : 0

Label : 0

6 - Conclusions



Best accuracy : RoBERTa 85%

Fast and great accuracy :
LR TF 2-gram + TF-IDF 4-gram + TF-IDF + Vader Score 82%

Still issue to predict the neutral

Future work :

- Acronym
- Word Embeddings