

# BUSINESS PROBLEM:

A SOCIAL MEDIA COMPANY WANTS TO IMPROVE USER ENGAGEMENT ON THEIR PLATFORM BY BETTER UNDERSTANDING THE EMOTIONS THEIR USERS ARE EXPRESSING IN THEIR POSTS USING A TEXT BASED RECOGNITION SYSTEM

## OBJECTIVES

Clean, parse and analyze textual data from twitter to predict emotions using Natural Language Processing

Create and use multiple classification models to figure out which models produces the best FI Score

Aim to lessen the amount of misclassification (misclassifying an emotion and failing to identify a certain emotion)



# DATA SOURCE:

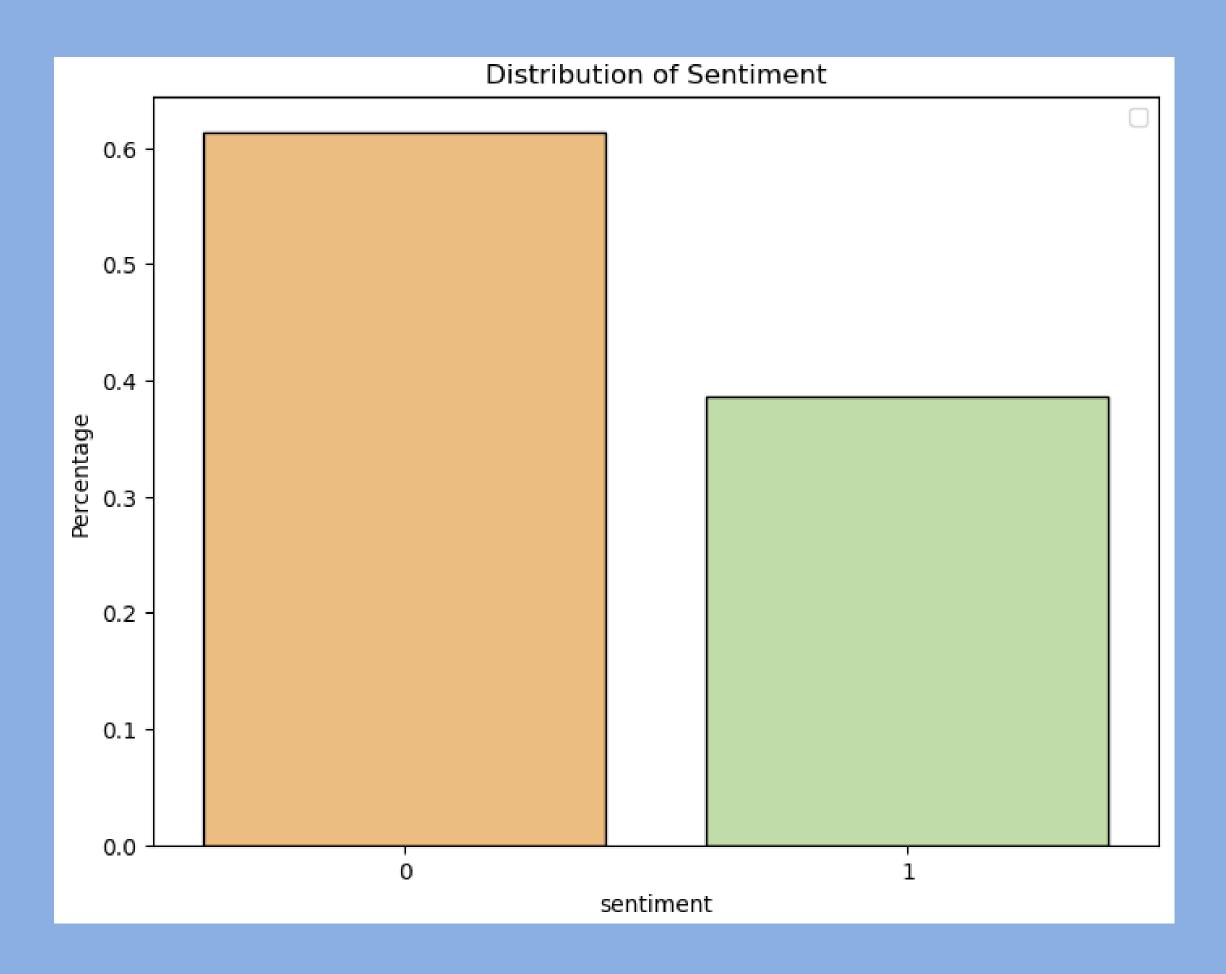


### **Emotion Detection from Text**

Predict emotion from textual data: Multi-class text classification

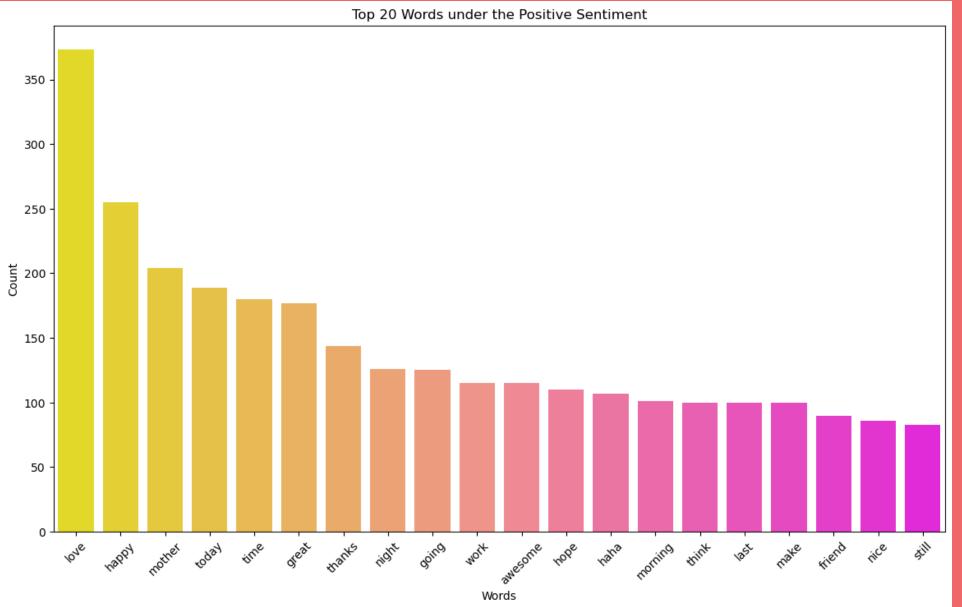
k kaggle.com

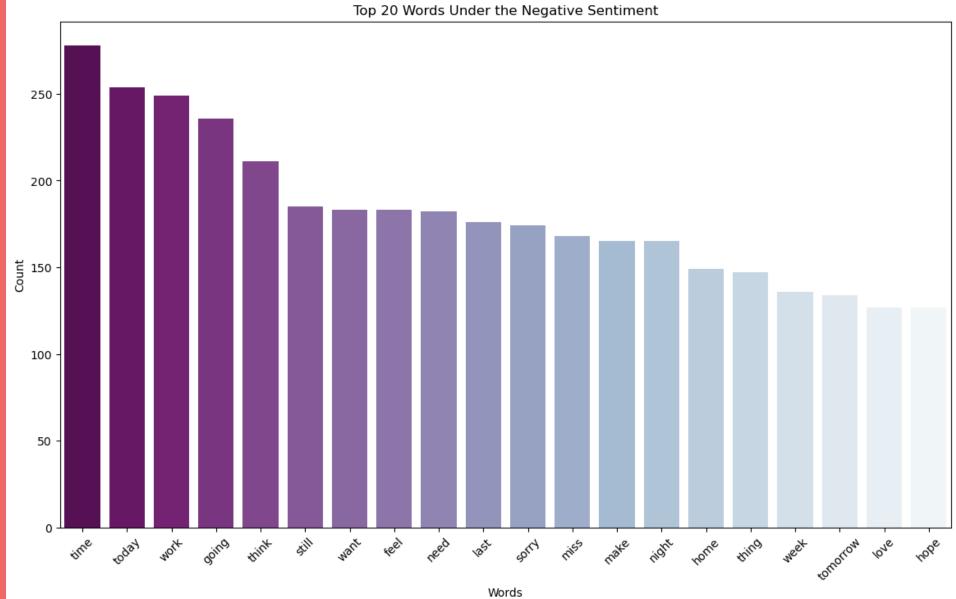




# THERE'S A CLASS IMBALANCE BETWEEN THE POSITIVE (1) AND NEGATIVE (0) CLASS

CLASS 0: 60% CLASS 1: 40%

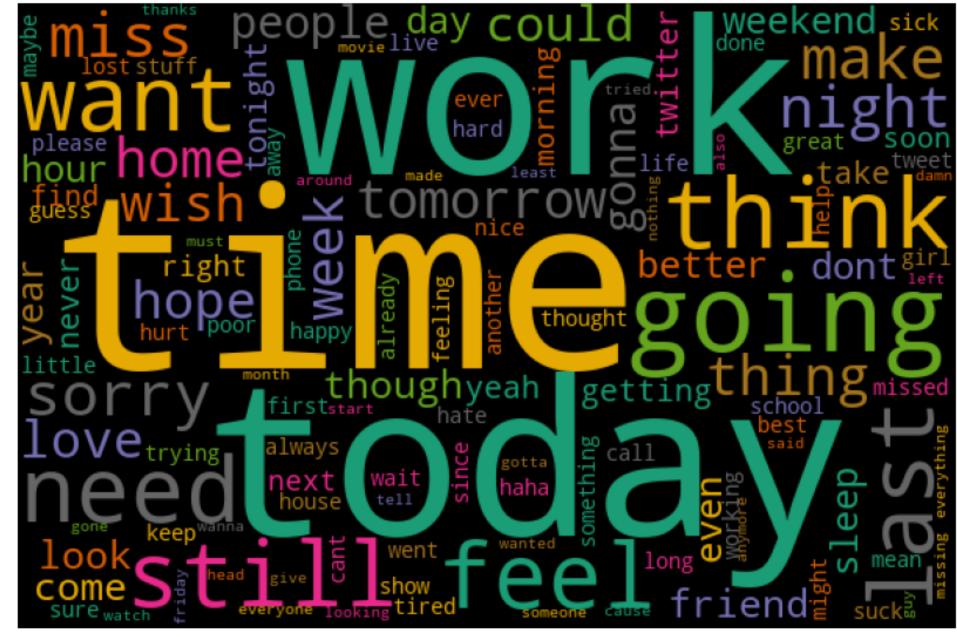










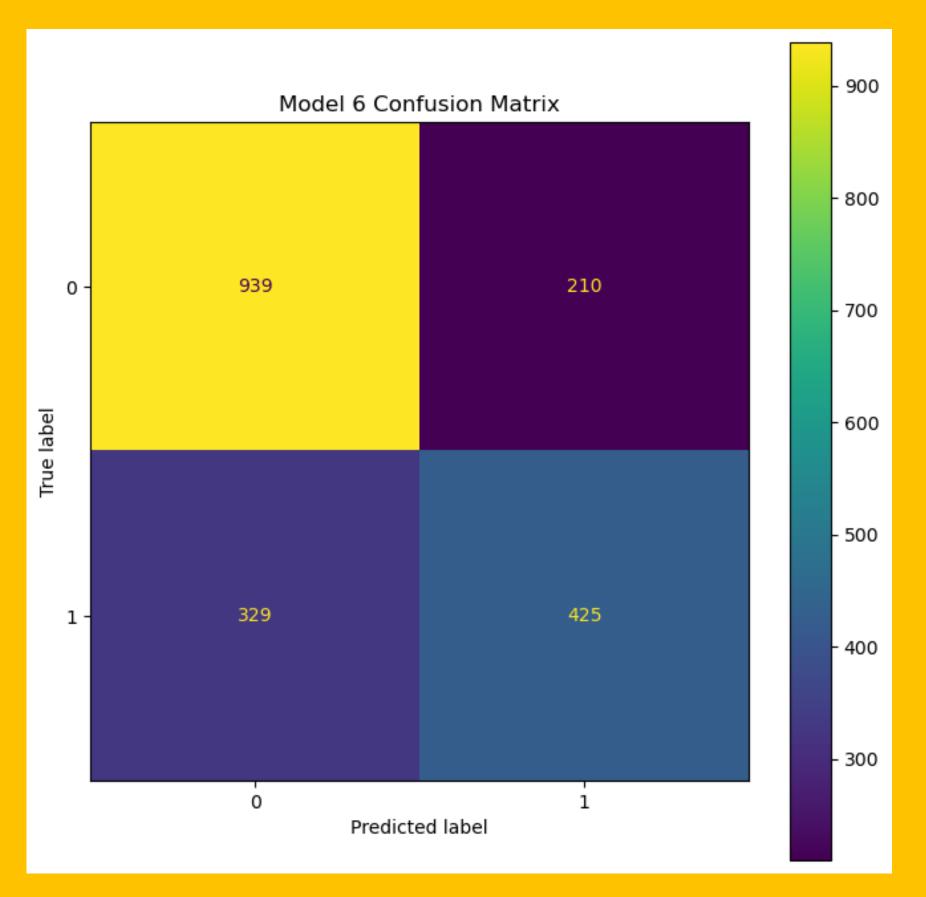


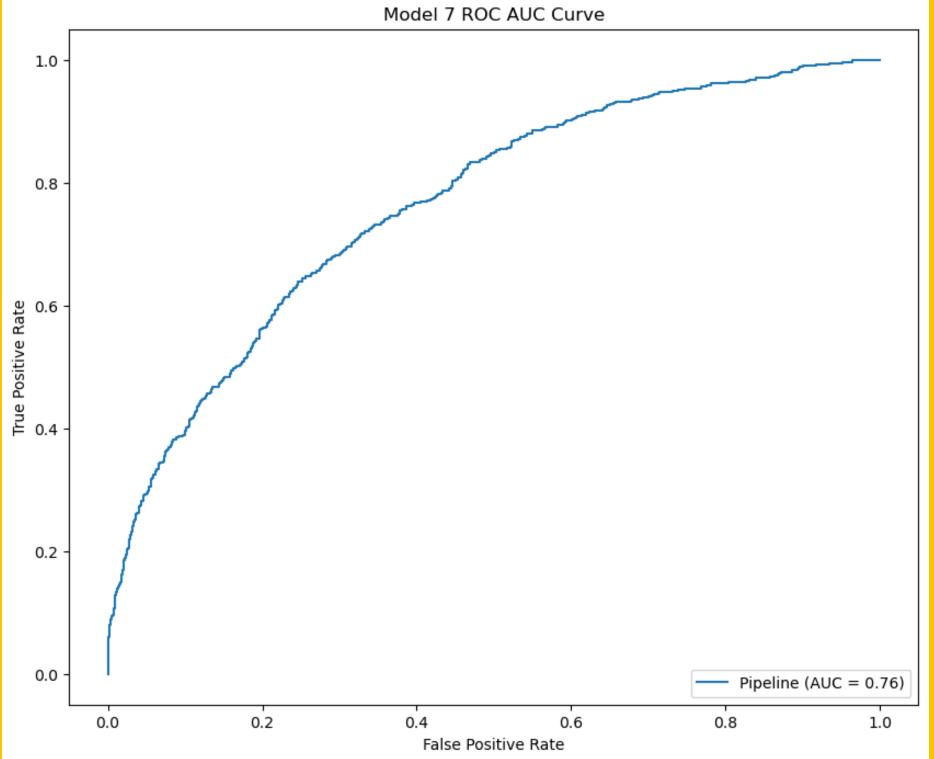
## BEST MODELS



	PRECISION	RECALL	F1 SCORE
STACKED CLASSIFIER (MULTINOMIAL NB AND XGBOOST)	CLASS 0: 0.74 CLASS 1: 0.67	CLASS 0: 0.82 CLASS 1: 0.56	CLASS 0: 0.78 CLASS 1: 0.62
RANDOM FOREST	CLASS 0: 0.76	CLASS 0: 0.72	CLASS 0: 0.74
	CLASS 1: 0.63	CLASS 1: 0.65	CLASS 1: 0.63
MULTIONOMIAL NAIVE BAYES	CLASS 0: 0.74	CLASS 0: 0.79	CLASS 0: 0.77
	CLASS 1: 0.65	CLASS 1: 0.58	CLASS 1: 0.61







## RECOMMENDATIONS

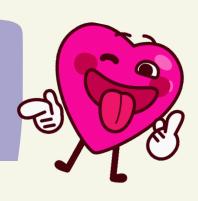


GATHER MORE PROPERLY CLASSIFIED DATA



WITH MORE PROPERLY CLASSIFIED TARGETS, WE CAN AIM TO LOWER BOTH FALSE NEGATIVES AND FALSE POSITIVES TO AVOID MISCLASSIFICATION

TRY USING BIGRAMS/TRIGRAMS TO CAPTURE MORE CONTEXTUAL INFORMATION THAN INDIVIDUAL WORDS







jocelarcelona30@gmail.com



www.linkedin.com/in/jocel-arcelona



https://github.com/JocelArcelona

