

```
In [1]: # Loading Dependencies
from path import Path
import pandas as pd
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

#load VADER
analyzer = SentimentIntensityAnalyzer()
```

```
In [2]: # Create SQL engine connection string; keeping in memory for validation, but
# to file ('sqlite:///reddit.db') if deployed

from sqlalchemy import create_engine
engine = create_engine('sqlite://', echo=False)
```

```
In [3]: # Loading Data
data = Path('Resources/reddit.csv')
reddit_df = pd.read_csv(data)
reddit_df.head()
```

Out[3]:

	title	score	subreddit	url	num_comments
0	UPVOTE so everyone sees we got SUPPORT	265029	wallstreetbets	https://i.redd.it/sgoqy8nyt2e61.png	1182
1	GME YOLO update — Jan 28 2021	230844	wallstreetbets	https://i.redd.it/opzucppb15e61.png	2353
2	CLASS ACTION AGAINST ROBINHOOD. Allowing peopl...	204920	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comment...	1831
3	GME YOLO update — Jan 27 2021 -----	185949	wallstreetbets	https://i.redd.it/a309gkm5yxd61.png	1549
4	Can we all take a moment and appreciate the Mo...	184517	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comment...	710

```
In [4]: # Checking DTypes
reddit_df.dtypes
```

```
Out[4]: title           object
score           int64
subreddit       object
url            object
num_comments    int64
body           object
date           object
dtype: object
```

```
In [5]: # Look at subreddit value counts
subreddit = reddit_df.subreddit.value_counts()
subreddit
```

```
Out[5]: investing      987
stocks              985
wallstreetbets      953
Name: subreddit, dtype: int64
```

```
In [6]: # Look at body value counts
body = reddit_df.body.value_counts()
body
```

```
Out[6]: Please use this thread to discuss your portfolio, learn of other stock ti
ckers, and help out users by giving constructive criticism.\n\nWhy quarte
rly? Public companies report earnings quarterly; many investors take thi
s as an opportunity to rebalance their portfolios. We highly recommend y
ou do some reading: A list of [relevant posts & book recommendations.](h
ttps://www.reddit.com/r/stocks/wiki/index#wiki_relevant_posts_.26amp.3B_b
ook_recommendations)\n\nYou can find stocks on your own by using a scanne
r like your broker's or [Finviz.](https://finviz.com/screener.ashx) To h
elp further, here's a list of [relevant websites.](https://www.reddit.co
m/r/stocks/wiki/index#wiki_relevant_websites.2Fapps)\n\nIf you don't have
a broker yet, see our [list of brokers](https://www.reddit.com/r/stocks/w
iki/index#wiki_brokers_for_investing) or search old posts. If you have
n't started investing or trading yet, then setup your [paper trading.](ht
tps://www.reddit.com/r/stocks/wiki/index#wiki_is_there_a_way_to_practice.
3F)\n\nBe aware of [Business Cycle Investing](https://ereseach.fidelity.
com/ereseach/markets_sectors/sectors/si_business_cycle.jhtml?tab=sibusin
ess) which Fidelity issues updates to the state of global business cycles
every 1 to 3 months (note: Fidelity changes their links often, so search
for it since their take on it is enlightening). [Investopedia's take on
the Business Cycle](https://www.investopedia.com/articles/financial/06121
```

```
In [7]: # Read reddit_df dataframe to new SQL database. Currently set to replace on
# we want it to append with new CSV info in the future.
```

```
reddit_df.to_sql('reddit_raw_db', con=engine, if_exists='replace')
engine.execute("SELECT * FROM reddit_raw_db").fetchmany(size=20)
```

```
Out[7]: [(0, 'UPVOTE so everyone sees we got SUPPORT', 265029, 'wallstreetbets',
'https://i.redd.it/sgoqy8nyt2e61.png', 11825, None, '2021-01-29 00:40:34'),
(1, 'GME YOLO update - Jan 28 2021', 230844, 'wallstreetbets', 'https://i.redd.it/opzucppb15e61.png', 23532, None, '2021-01-29 08:06:23'),
(2, 'CLASS ACTION AGAINST ROBINHOOD. Allowing people to only sell is the definition of market manipulation. A class action must be started, Robinhood has made plenty of money off selling info about our trades to the hedge funds to be able to pay out a little for causing people to loose money now', 204920, 'wallstreetbets', 'https://www.reddit.com/r/wallstreetbets/comments/l6x130/class_action_against_robinhood_allowing_people_to/', 18318, 'LEAVE ROBINHOOD. They dont deserve to make money off us after the millions they caused in losses. It might take a couple of days, but send Robinhood to the ground and GME to the moon.', '2021-01-29 00:49:11'),
(3, 'GME YOLO update - Jan 27 2021 -----
-- guess i need 102 characters in title now', 185949, 'wallstreetbets', 'https://i.redd.it/a309gkm5yxd61.png', 15495, None, '2021-01-28 08:15:35'),
(4, 'Can we all take a moment and appreciate the Mods who have grinded through this and held our community together. Thank you.', 184517, 'wallstreetbets', 'https://www.reddit.com/r/wallstreetbets/comments/l6jobf/can_we_all_take_a_moment_and_appreciate_the_mods/', 7105, None, '2021-01-28 11:57:32'),
(5, "IT'S POWER TO THE TRADERS NOW", 180878, 'wallstreetbets', 'https://v.redd.it/lu8aekujd6e61', 5338, None, '2021-01-29 12:35:58'),
(6, "Crazy mannnnnnn. We can't let this slide at all", 166492, 'wallstreetbets', 'https://i.redd.it/8bels7zob4e61.jpg', 2485, None, '2021-01-29 05:42:01'),
(7, 'It runs very deep, my friends.', 165358, 'wallstreetbets', 'https://i.redd.it/b8zy92sy85e61.jpg', 2796, None, '2021-01-29 08:47:55'),
(8, 'AOC got our backs', 164232, 'wallstreetbets', 'https://i.redd.it/3v5c5quc5yd61.png', 5394, None, '2021-01-28 08:55:22'),
(9, 'I'm so proud of you all.', 160021, 'wallstreetbets', 'https://i.redd.it/ykiux57d65e61.jpg', 2536, None, '2021-01-29 08:33:21'),
(10, "AMERICANS CAN'T BUY GME, BB?!?! HELP IS ON THE WAY!! CANADIANS!! WE HAVE NO SUCH RESTRICTIONS!! BUY ALL THE GME AND BB YOU CAN!!!", 157980, 'wallstreetbets', 'https://www.reddit.com/r/wallstreetbets/comments/l6y25u/americans_cant_buy_gme_bb_help_is_on_the_way/', 23210, 'Edit* Not just Canadians!! The whole world is coming to help now!!\n\nEdit2* WE ARE A GLOBAL FORCE TO BE RECKONED WITH!!! HOLD MY BROTHERS AND SISTERS!', '2021-01-29 01:32:00'),
(11, 'Congress might do something for once', 157535, 'wallstreetbets', 'https://i.redd.it/3lo7rt55p5e61.jpg', 6969, None, '2021-01-29 10:18:35'),
(12, 'IMPORTANT!! THERE ARE NO SELLERS!!! DO NOT PANIC IF YOU SEE PRICE DROP HOLD THE LINE! UPVOTE SO PEOPLE DON'T PANIC.', 148351, 'wallstreetbets', 'https://i.redd.it/j6glpaf4h3e61.jpg', 5263, None, '2021-01-29 02:50:05'),
(13, 'Not sure if anyone posted this here, but yikes', 145918, 'wallstreetbets', 'https://i.redd.it/76bcnhs2k5e61.jpg', 5871, None, '2021-01-29 09:50:11'),
```

```
(14, 'Doing my part', 139403, 'wallstreetbets', 'https://i.redd.it/af2t6dp725e61.png', 2870, None, '2021-01-29 08:10:30'),
(15, 'An Open Letter to Melvin Capital, CNBC, Boomers, and WSB', 138569, 'wallstreetbets', 'https://www.reddit.com/r/wallstreetbets/comments/l6omry/an_open_letter_to_melvin_capital_cnbc_boomers_and/', 7802, ' **Mods do not delete, this is important to me, please read**\n\n&#x200B;\n\nI was in my early teens during the \'08 crisis. I vividly remember the e ... (49 61 characters truncated) ... ods and they keep fucking removing the post. I have no idea how to get this to stick and its important to me that the people I\'m addressing read it.', '2021-01-28 16:06:14'),
(16, 'Mic drop', 137716, 'wallstreetbets', 'https://i.redd.it/ozzh6wv9d5e61.jpg', 1933, None, '2021-01-29 09:12:04'),
(17, "YOU RETARDS ARE ACTUALLY HOLDING THE LINE!!!! DON'T STOP 💎💎💎💎 HANDS!!!!", 136858, 'wallstreetbets', 'https://i.redd.it/k7t143dkn4e61.jpg', 10018, None, '2021-01-29 06:48:39'),
(18, 'Robinhood will not let new options or shares be bought apparently. This is market manipulation lol', 136707, 'wallstreetbets', 'https://i.redd.it/vboawqu6o2e61.jpg', 16053, None, '2021-01-29 00:07:54'),
(19, 'If I can hold at -$10MM then you can hold too. 💎👉', 136279, 'wallstreetbets', 'https://i.redd.it/e2ra46rsk3e61.jpg', 7318, None, '2021-01-29 03:11:24')]
```

```
In [8]: # Remove the 'body' column.
reddit_df.drop(['body'], axis=1, inplace=True)
reddit_df.head()
```

Out[8]:

	title	score	subreddit	url	num_comments
0	UPVOTE so everyone sees we got SUPPORT	265029	wallstreetbets	https://i.redd.it/sgoqy8nyt2e61.png	1182
1	GME YOLO update — Jan 28 2021	230844	wallstreetbets	https://i.redd.it/opzucppb15e61.png	2353
2	CLASS ACTION AGAINST ROBINHOOD. Allowing people...	204920	wallstreetbets	https://www.reddit.com/r/wallstreetbets/commen...	1831
3	GME YOLO update — Jan 27 2021 -----	185949	wallstreetbets	https://i.redd.it/a309gkm5yxd61.png	1549
4	Can we all take a moment and appreciate the Mo...	184517	wallstreetbets	https://www.reddit.com/r/wallstreetbets/commen...	710

```
In [9]: # Remove rows that have at least 1 null value.
reddit_df.dropna()
```

Out[9]:

	title	score	subreddit	url	num_con
0	UPVOTE so everyone sees we got SUPPORT	265029	wallstreetbets	https://i.redd.it/sgoqy8nyt2e61.png	
1	GME YOLO update — Jan 28 2021	230844	wallstreetbets	https://i.redd.it/opzucppb15e61.png	
2	CLASS ACTION AGAINST ROBINHOOD. Allowing peopl...	204920	wallstreetbets	https://www.reddit.com/r/wallstreetbets/commen...	
3	GME YOLO update — Jan 27 2021 -----	185949	wallstreetbets	https://i.redd.it/a309gkm5yxd61.png	
4	Can we all take a moment and appreciate the Mo...	184517	wallstreetbets	https://www.reddit.com/r/wallstreetbets/commen...	
...
2920	DID WE MISS THE BOTTOM?! How are people this i...	348	stocks	https://www.reddit.com/r/stocks/comments/g1m6u...	
2921	Favorite Solar Stock(s)?	348	stocks	https://www.reddit.com/r/stocks/comments/jmct3...	
2922	"NIO forms battery asset company"	352	stocks	https://www.reddit.com/r/stocks/comments/id8z3...	
2923	Amazon is building a \$1.5 billion hub for its ...	349	stocks	https://www.reddit.com/r/stocks/comments/5rghg...	
2924	Weed legalized in Canada!	356	stocks	https://www.reddit.com/r/stocks/comments/8skaf...	

2925 rows × 6 columns

```
In [10]: # Checking Sentiment Scores
def sentiment_analyzer_scores(sentence):
    score = analyzer.polarity_scores(sentence)
    print("{:-<40} {}".format(sentence, str(score)))

sentiment_analyzer_scores('UPVOTE so everyone sees we got SUPPORT')
```

```
UPVOTE so everyone sees we got SUPPORT-- {'neg': 0.0, 'neu': 0.636, 'pos': 0.364, 'compound': 0.5319}
```

```
In [11]: #Add VADER metrics to dataframe
reddit_df['compound'] = [analyzer.polarity_scores(v)['compound'] for v in reddit_df['text']]
reddit_df['neg'] = [analyzer.polarity_scores(v)['neg'] for v in reddit_df['text']]
reddit_df['neu'] = [analyzer.polarity_scores(v)['neu'] for v in reddit_df['text']]
reddit_df['pos'] = [analyzer.polarity_scores(v)['pos'] for v in reddit_df['text']]
reddit_df.head()
```

Out[11]:

	title	score	subreddit	url	num_comments
0	UPVOTE so everyone sees we got SUPPORT	265029	wallstreetbets	https://i.redd.it/sgoqy8nyt2e61.png	1182
1	GME YOLO update — Jan 28 2021	230844	wallstreetbets	https://i.redd.it/opzucppb15e61.png	2353
2	CLASS ACTION AGAINST ROBINHOOD. Allowing people...	204920	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comments/204920/class_action_against_robinhood_allowing_people_to/	1831
3	GME YOLO update — Jan 27 2021 -----...	185949	wallstreetbets	https://i.redd.it/a309gkm5yxd61.png	1549
4	Can we all take a moment and appreciate the Mo...	184517	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comments/184517/can_we_all_take_a_moment_and_appreciate_the_moment/	710

```
In [12]: reddit_df.to_csv('Resources/reddit2.csv')
```

```
In [13]: reddit_groups = reddit_df.groupby("subreddit")
```

```
In [14]: # Grouping Vader Scores for each Subreddit
reddit_groups.mean()
```

Out[14]:

	score	num_comments	compound	neg	neu	pos
subreddit						
investing	1779.078014	424.946302	-0.009373	0.080139	0.842621	0.077234
stocks	1405.035533	311.643655	0.038948	0.060024	0.852797	0.087179
wallstreetbets	29348.439664	2753.270724	0.046143	0.066779	0.834423	0.098805

```
In [15]: reddit_df = reddit_df[reddit_df["subreddit"]!="stocks"]
reddit_df.tail()
```

Out[15]:

	title	score	subreddit	url	num_cc
1935	does anyone here have friends and family still...	898	investing	https://www.reddit.com/r/investing/comments/ef...	
1936	Jobs growth soars in November as payrolls surg...	905	investing	https://www.reddit.com/r/investing/comments/e6...	
1937	Yale economists argue that "the most financial...	894	investing	https://www.reddit.com/r/investing/comments/en...	
1938	Amazon earnings beat: 6.04pershare, vs.5....	897	investing	https://www.reddit.com/r/investing/comments/al...	
1939	U.S. stocks plunge after report that former na...	890	investing	https://www.reddit.com/r/investing/comments/7g...	

```
In [16]: # Changing Subreddit values to be 1 if from WSB and 0 if not.
reddit_df['subreddit'] = reddit_df['subreddit'].apply(lambda x:1 if x == "w
/Users/brendandevidal/opt/anaconda3/envs/PythonData/lib/python3.7/site-pa
ckages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
```

```
In [17]: reddit_df.head()
```

```
Out[17]:
```

	title	score	subreddit	url	num_comments
0	UPVOTE so everyone sees we got SUPPORT	265029	1	https://i.redd.it/sgoqy8nyt2e61.png	11825
1	GME YOLO update — Jan 28 2021	230844	1	https://i.redd.it/opzucppb15e61.png	23532
2	CLASS ACTION AGAINST ROBINHOOD. Allowing peopl...	204920	1	https://www.reddit.com/r/wallstreetbets/comments/...	18318
3	GME YOLO update — Jan 27 2021 -----... -----...	185949	1	https://i.redd.it/a309gkm5yxd61.png	15495
4	Can we all take a moment and appreciate the Mo...	184517	1	https://www.reddit.com/r/wallstreetbets/comments/...	7105

```
In [18]: reddit_df.tail()
```

```
Out[18]:
```

	title	score	subreddit	url	num_cc
1935	does anyone here have friends and family still...	898	0	https://www.reddit.com/r/investing/comments/ef...	
1936	Jobs growth soars in November as payrolls surg...	905	0	https://www.reddit.com/r/investing/comments/e6...	
1937	Yale economists argue that "the most financial...	894	0	https://www.reddit.com/r/investing/comments/en...	
1938	Amazon earnings beat: 6.04pershare, vs.5....	897	0	https://www.reddit.com/r/investing/comments/al...	
1939	U.S. stocks plunge after report that former na...	890	0	https://www.reddit.com/r/investing/comments/7g...	


```
In [19]: # Read the cleaned and VADER modified reddit_df dataframe to new SQL databa
# until we can decide how we want it to append with new CSV info in the fut
```

```
reddit_df.to_sql('reddit_cleaned_with_VADER_db', con=engine, if_exists='rep
engine.execute("SELECT * FROM reddit_cleaned_with_VADER_db").fetchmany(size
```

```
Out[19]: [(0, 'UPVOTE so everyone sees we got SUPPORT', 265029, 1, 'https://i.red
d.it/sgoqy8nyt2e61.png', 11825, '2021-01-29 00:40:34', 0.5319, 0.0, 0.63
6, 0.364),
(1, 'GME YOLO update - Jan 28 2021', 230844, 1, 'https://i.redd.it/opzuc
ppb15e61.png', 23532, '2021-01-29 08:06:23', 0.4278, 0.0, 0.679, 0.321),
(2, 'CLASS ACTION AGAINST ROBINHOOD. Allowing people to only sell is the
definition of market manipulation. A class action must be started, Robinh
ood has made plenty of money off selling info about our trades to the hed
ge funds to be able to pay out a little for causing people to loose money
now', 204920, 1, 'https://www.reddit.com/r/wallstreetbets/comments/l6x13
0/class_action_against_robinhood_allowing_people_to/', 18318, '2021-01-29
00:49:11', -0.5994, 0.107, 0.893, 0.0),
(3, 'GME YOLO update - Jan 27 2021 -----
-- guess i need 102 characters in title now', 185949, 1, 'https://i.redd.
it/a309gkm5yxd61.png', 15495, '2021-01-28 08:15:35', 0.4278, 0.0, 0.841,
0.159),
(4, 'Can we all take a moment and appreciate the Mods who have grinded t
hrough this and held our community together. Thank you.', 184517, 1, 'htt
ps://www.reddit.com/r/wallstreetbets/comments/l6jobf/can_we_all_take_a_mo
ment_and_appreciate_the_mods/', 7105, '2021-01-28 11:57:32', 0.6369, 0.0,
0.794, 0.206),
(5, "IT'S POWER TO THE TRADERS NOW", 180878, 1, 'https://v.redd.it/lu8ae
kujd6e61', 5338, '2021-01-29 12:35:58', 0.0, 0.0, 1.0, 0.0),
(6, "Crazy mannnnnnn. We can't let this slide at all", 166492, 1, 'http
s://i.redd.it/8bels7zob4e61.jpg', 2485, '2021-01-29 05:42:01', -0.34, 0.2
31, 0.769, 0.0),
(7, 'It runs very deep, my friends.', 165358, 1, 'https://i.redd.it/b8zy
92sy85e61.jpg', 2796, '2021-01-29 08:47:55', 0.5209, 0.0, 0.598, 0.402),
(8, 'AOC got our backs', 164232, 1, 'https://i.redd.it/3v5c5quc5yd61.pn
g', 5394, '2021-01-28 08:55:22', -0.0516, 0.286, 0.714, 0.0),
(9, 'I'm so proud of you all.', 160021, 1, 'https://i.redd.it/ykiux57d65
e61.jpg', 2536, '2021-01-29 08:33:21', 0.5256, 0.0, 0.596, 0.404),
(10, "AMERICANS CAN'T BUY GME, BB?!?! HELP IS ON THE WAY!! CANADIANS!! W
E HAVE NO SUCH RESTRICTIONS!! BUY ALL THE GME AND BB YOU CAN!!!", 157980,
1, 'https://www.reddit.com/r/wallstreetbets/comments/l6y25u/americans_can
t_buy_gme_bb_help_is_on_the_way/', 23210, '2021-01-29 01:32:00', 0.4639,
0.077, 0.774, 0.149),
(11, 'Congress might do something for once', 157535, 1, 'https://i.redd.
it/3lo7rt55p5e61.jpg', 6969, '2021-01-29 10:18:35', 0.0, 0.0, 1.0, 0.0),
(12, 'IMPORTANT!! THERE ARE NO SELLERS!!! DO NOT PANIC IF YOU SEE PRICE
DROP HOLD THE LINE! UPVOTE SO PEOPLE DON'T PANIC.', 148351, 1, 'https://
i.redd.it/j6glpaf4h3e61.jpg', 5263, '2021-01-29 02:50:05', -0.6736, 0.30
6, 0.542, 0.152),
(13, 'Not sure if anyone posted this here, but yikes', 145918, 1, 'http
s://i.redd.it/76bcnhs2k5e61.jpg', 5871, '2021-01-29 09:50:11', -0.1232,
0.156, 0.844, 0.0),
(14, 'Doing my part', 139403, 1, 'https://i.redd.it/af2t6dp725e61.png',
2870, '2021-01-29 08:10:30', 0.0, 0.0, 1.0, 0.0),
(15, 'An Open Letter to Melvin Capital, CNBC, Boomers, and WSB', 138569,
1, 'https://www.reddit.com/r/wallstreetbets/comments/l6omry/an_open_lette
```

```

r_to_melvin_capital_cnn_boomers_and/', 7802, '2021-01-28 16:06:14', 0.0,
0.0, 1.0, 0.0),
(16, 'Mic drop', 137716, 1, 'https://i.redd.it/ozzh6wv9d5e61.jpg', 1933,
'2021-01-29 09:12:04', -0.2732, 0.677, 0.323, 0.0),
(17, "YOU RETARDS ARE ACTUALLY HOLDING THE LINE!!!! DON'T STOP 💎💎💎💎
HANDS!!!!", 136858, 1, 'https://i.redd.it/k7t143dkn4e61.jpg', 10018, '202
1-01-29 06:48:39', 0.5571, 0.0, 0.825, 0.175),
(18, 'Robinhood will not let new options or shares be bought apparently.
This is market manipulation lol', 136707, 1, 'https://i.redd.it/vboawqu6o
2e61.jpg', 16053, '2021-01-29 00:07:54', 0.4215, 0.109, 0.644, 0.248),
(19, 'If I can hold at -$10MM then you can hold too. 💎🙌', 136279, 1,
'https://i.redd.it/e2ra46rsk3e61.jpg', 7318, '2021-01-29 03:11:24', 0.0,
0.0, 1.0, 0.0)]

```

```
In [20]: # Read from DB into a dataframe for ML model
```

```
reddit_cleaned_with_VADER_df = pd.read_sql("reddit_cleaned_with_VADER_db",
```

```
In [21]: reddit_cleaned_with_VADER_df.head()
```

```
Out[21]:
```

	index	title	score	subreddit	url	num_comr
0	0	UPVOTE so everyone sees we got SUPPORT	265029	1	https://i.redd.it/sgoqy8nyt2e61.png	
1	1	GME YOLO update — Jan 28 2021	230844	1	https://i.redd.it/opzucppb15e61.png	:
2	2	CLASS ACTION AGAINST ROBINHOOD. Allowing peopl...	204920	1	https://www.reddit.com/r/wallstreetbets/commen...	
3	3	GME YOLO update — Jan 27 2021 -----	185949	1	https://i.redd.it/a309gkm5yxd61.png	
4	4	Can we all take a moment and appreciate the Mo...	184517	1	https://www.reddit.com/r/wallstreetbets/commen...	

In [22]: *# combined_db with JOIN via SQL*

```
execute("DROP TABLE IF EXISTS combined_db")
execute("CREATE TABLE combined_db AS SELECT * FROM reddit_raw_db JOIN reddit
execute("SELECT * FROM combined_db").fetchmany(size=20)
```

Out[22]: [(0, 'UPVOTE so everyone sees we got SUPPORT', 265029, 'wallstreetbets',
'https://i.redd.it/sgoqy8nyt2e61.png', 11825, None, '2021-01-29 00:40:34', 0, 'UPVOTE so everyone sees we got SUPPORT', 265029, 1, 'https://i.redd.it/sgoqy8nyt2e61.png', 11825, '2021-01-29 00:40:34', 0.5319, 0.0, 0.636, 0.364),
(1, 'GME YOLO update - Jan 28 2021', 230844, 'wallstreetbets', 'https://i.redd.it/opzucppb15e61.png', 23532, None, '2021-01-29 08:06:23', 1, 'GME YOLO update - Jan 28 2021', 230844, 1, 'https://i.redd.it/opzucppb15e61.png', 23532, '2021-01-29 08:06:23', 0.4278, 0.0, 0.679, 0.321),
(2, 'CLASS ACTION AGAINST ROBINHOOD. Allowing people to only sell is the definition of market manipulation. A class action must be started, Robinhood has made plenty of money off selling info about our trades to the hedge funds to be able to pay out a little for causing people to loose money now', 204920, 'wallstreetbets', 'https://www.reddit.com/r/wallstreetbets/comments/l6x130/class_action_against_robinhood_allowing_people_to/', 18318, 'LEAVE ROBINHOOD. They dont deserve to make money off us after the millions they caused in losses. It might take a couple of days, but send Robinhood to the ground and GME to the moon.', '2021-01-29 00:49:11', 2, 'CLASS ACTION AGAINST ROBINHOOD. Allowing people to only sell is the definit

In [23]: *# Creating X and Y sets*

```
y = reddit_cleaned_with_VADER_df["subreddit"]
X = reddit_cleaned_with_VADER_df.drop(columns=["index", "subreddit", "title",
```

In [24]: *# Breaking sets into train and test.*

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,
                                                    y,
                                                    random_state=1,
                                                    stratify=y)

X_train.shape
```

Out[24]: (1455, 4)

In [25]: y.value_counts()

Out[25]: 0 987
1 953
Name: subreddit, dtype: int64

```
In [26]: # NOT NEEDED WHEN TAKING OUT SCORE AND NUM_COMMENTS
# create scaler instances
#scaler = StandardScaler()

# fit scaler
#X_scaler = scaler.fit(X_train)

# scale data
#X_train_scaled = X_scaler.transform(X_train)
#X_test_scaled = X_scaler.transform(X_test)
```

```
In [27]: from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(solver='lbfgs',
                                max_iter=200,
                                random_state=1)
```

```
In [28]: classifier.fit(X_train, y_train)
#classifier.fit(X_train_scaled, y_train)
```

```
Out[28]: LogisticRegression(max_iter=200, random_state=1)
```

```
In [29]: y_pred = classifier.predict(X_test)
results = pd.DataFrame({"Prediction": y_pred, "Actual": y_test}).reset_index()
results.head(20)
```

8	1	1
9	0	1
10	1	1
11	0	0
12	1	1
13	1	0
14	1	0
15	1	1
16	0	1
17	1	0
18	1	1

```
In [30]: from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, y_pred))
```

```
0.5092783505154639
```

Optimization Attempt #1 - Add Additional Hidden Layers

```
In [31]: #X_train_scaled.shape[1]
X_train.shape[1]
```

Out[31]: 4

```
In [32]: # define model
number_input_features = X_train.shape[1]
#number_input_features = X_train_scaled.shape[1]
hidden_nodes_layer1 = 80
hidden_nodes_layer2 = 30
hidden_nodes_layer3 = 20

nn = tf.keras.models.Sequential()

# first hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer1, input_dim=number_in

# second hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer2, activation='relu'))

#third hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer3, activation='relu'))

# output layer
nn.add(tf.keras.layers.Dense(units=1, activation='sigmoid'))

# check structure of model
nn.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 80)	400
dense_1 (Dense)	(None, 30)	2430
dense_2 (Dense)	(None, 20)	620
dense_3 (Dense)	(None, 1)	21
Total params: 3,471		
Trainable params: 3,471		
Non-trainable params: 0		

```
In [33]: # Compile the model
nn.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy
```

```
In [34]: # Train the model
fit_model = nn.fit(X_train, y_train, epochs=100)
#fit_model = nn.fit(X_train_scaled, y_train, epochs=100)

Epoch 1/100
46/46 [=====] - 0s 795us/step - loss: 0.6738 - a
ccuracy: 0.6284
Epoch 5/100
46/46 [=====] - 0s 824us/step - loss: 0.6634 - a
ccuracy: 0.6453
Epoch 6/100
46/46 [=====] - 0s 901us/step - loss: 0.6587 - a
ccuracy: 0.6448
Epoch 7/100
46/46 [=====] - 0s 858us/step - loss: 0.6539 - a
ccuracy: 0.6536
Epoch 8/100
46/46 [=====] - 0s 919us/step - loss: 0.6353 - a
ccuracy: 0.6700
Epoch 9/100
46/46 [=====] - 0s 2ms/step - loss: 0.6467 - acc
uracy: 0.6468
Epoch 10/100
46/46 [=====] - 0s 886us/step - loss: 0.6378 - a
ccuracy: 0.6634
```

```
In [35]: # Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(X_test,y_test,verbose=2)
#model_loss, model_accuracy = nn.evaluate(X_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")

16/16 - 0s - loss: 0.6389 - accuracy: 0.6165
Loss: 0.6389384865760803, Accuracy: 0.6164948344230652
```

Optimization Attempt #2: Adding Neurons in Layers & tanh for relu

```
In [36]: # define model
number_input_features = X_train.shape[1]
hidden_nodes_layer1 = 90
hidden_nodes_layer2 = 40
hidden_nodes_layer3 = 30

nn = tf.keras.models.Sequential()

# first hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer1, input_dim=number_in

# second hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer2, activation='tanh'))

#third hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer3, activation='tanh'))

# output layer
nn.add(tf.keras.layers.Dense(units=1, activation='sigmoid'))

# check structure of model
nn.summary()
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
=====	=====	=====
dense_4 (Dense)	(None, 90)	450
dense_5 (Dense)	(None, 40)	3640
dense_6 (Dense)	(None, 30)	1230
dense_7 (Dense)	(None, 1)	31
=====	=====	=====
Total params: 5,351		
Trainable params: 5,351		
Non-trainable params: 0		
=====		

```
In [37]: # Compile the model
nn.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy
```

```
In [38]: # Train the model
fit_model = nn.fit(X_train, y_train, epochs=63)
accuracy: 0.5511
Epoch 2/63
46/46 [=====] - 0s 828us/step - loss: 0.6925 - a
ccuracy: 0.4989
Epoch 3/63
46/46 [=====] - 0s 820us/step - loss: 0.6926 - a
ccuracy: 0.5212
Epoch 4/63
46/46 [=====] - 0s 796us/step - loss: 0.6872 - a
ccuracy: 0.5553
Epoch 5/63
46/46 [=====] - 0s 943us/step - loss: 0.6927 - a
ccuracy: 0.5125
Epoch 6/63
46/46 [=====] - 0s 914us/step - loss: 0.6907 - a
ccuracy: 0.5065
Epoch 7/63
46/46 [=====] - 0s 827us/step - loss: 0.6905 - a
ccuracy: 0.5159
Epoch 8/63
46/46 [=====] - 0s 701us/step - loss: 0.6860 - a
ccuracy: 0.5211
```

```
In [39]: # Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(X_test,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")

16/16 - 0s - loss: 0.6519 - accuracy: 0.6082
Loss: 0.6518993973731995, Accuracy: 0.6082473993301392
```

Optimization Attempt #3: Change Activation Function w/ Callback

```
In [40]: # define checkpoint path and filenames
import os
os.makedirs("checkpoints/",exist_ok=True)
checkpoint_path = "checkpoints/weights.{epoch:02d}.hdf5"
```



```

In [41]: # define model
number_input_features = X_train.shape[1]
hidden_nodes_layer1 = 80
hidden_nodes_layer2 = 30
hidden_nodes_layer3 = 20

nn = tf.keras.models.Sequential()

# first hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer1, input_dim=number_in

# second hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer2, activation='relu'))

# output layer
nn.add(tf.keras.layers.Dense(units=1, activation='tanh'))

# check structure of model
nn.summary()

```

Model: "sequential_2"

Layer (type)	Output Shape	Param #
=====		
dense_8 (Dense)	(None, 80)	400
dense_9 (Dense)	(None, 30)	2430
dense_10 (Dense)	(None, 1)	31
=====		

Total params: 2,861

Trainable params: 2,861

Non-trainable params: 0

```
In [42]: from tensorflow.keras.callbacks import ModelCheckpoint

# compile model
nn.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"])

# create callback that saves weights every 5 epochs
cp_callback = ModelCheckpoint(
    filepath=checkpoint_path,
    verbose=1,
    save_weights_only=True,
    save_freq=1000)

# train model
fit_model = nn.fit(X_train,y_train,epochs=100,callbacks=[cp_callback])

# evaluate model using test data
model_loss, model_accuracy = nn.evaluate(X_test,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")

Epoch 1/100
46/46 [=====] - 0s 2ms/step - loss: 0.8038 - accuracy: 0.5120
Epoch 2/100
46/46 [=====] - 0s 798us/step - loss: 0.6807 - accuracy: 0.5449
Epoch 3/100
46/46 [=====] - 0s 725us/step - loss: 0.6689 - accuracy: 0.5455
Epoch 4/100
46/46 [=====] - 0s 907us/step - loss: 0.6629 - accuracy: 0.6123
Epoch 5/100
46/46 [=====] - 0s 856us/step - loss: 0.6570 - accuracy: 0.6135
Epoch 6/100
46/46 [=====] - 0s 806us/step - loss: 0.6572 - accuracy: 0.6180
Epoch 7/100
46/46 [=====] - 0s 849us/step - loss: 0.6515 - a
```

In []:

In []: