

Abstract geometric lines in black on a white background, forming various overlapping polygons and shapes.

CAN SOCIAL MEDIA AFFECT A SHARE PRICE?

Brian Higgins | April 2022 | BrainStation

**GAME STOP SHARE PRICE
JANUARY 2021**

1800%

PROJECT OVERVIEW



PROBLEM

Can we use Reddit Comments to predict Game Stop's Share Price



MOTIVATION

In the future, be able to spot situations like with Game Stop

DATA COLLECTION

**Reddit Social
Media**



**Game Stop
Share Price**



DATA DESCRIPTION

REDDIT DATA

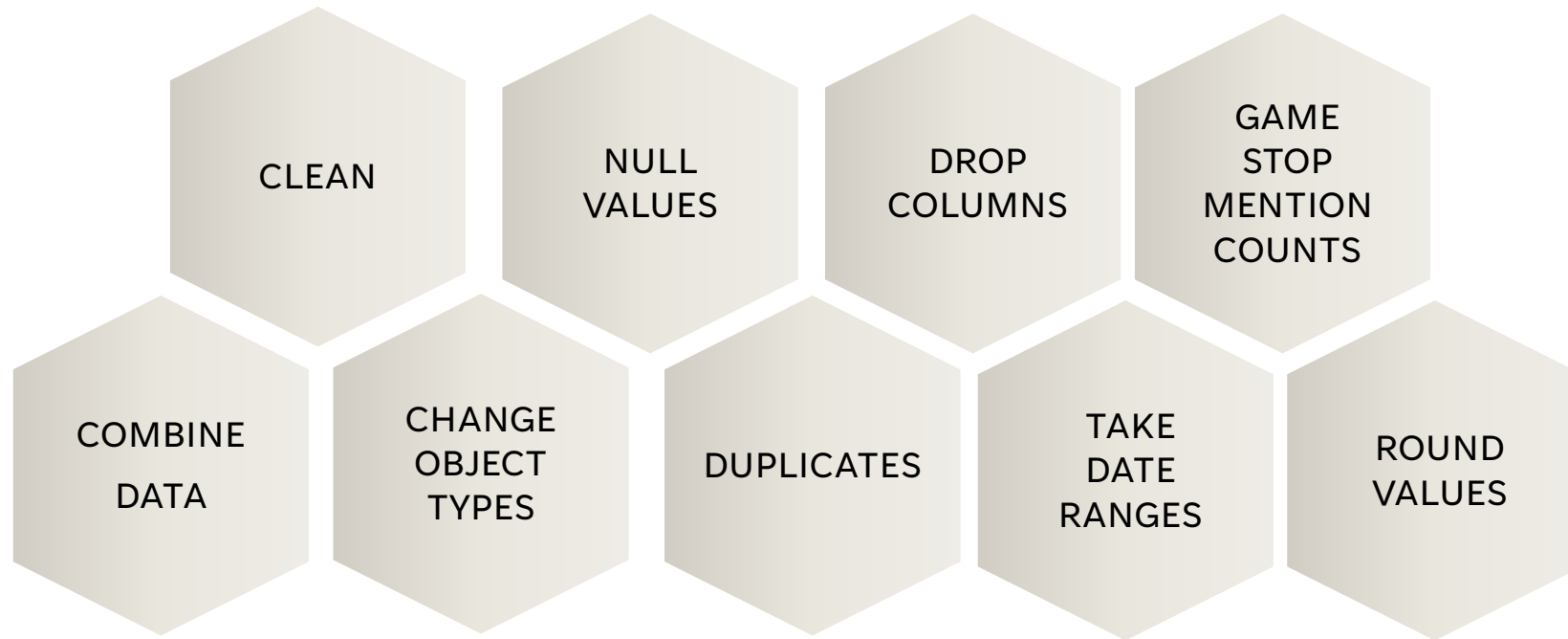
14 Financial Subreddits
2021
Social Media Comments

GAME STOP SHARE DATA

1 Data Set – 15 years
Share Price Information

1,500,994
rows

DATA PROCESSING



155,505 Game Stop
Mentions

FEATURES

Target Variable?
Share Price

Reddit Features

Share Price Features

New Features

Lag Ranges

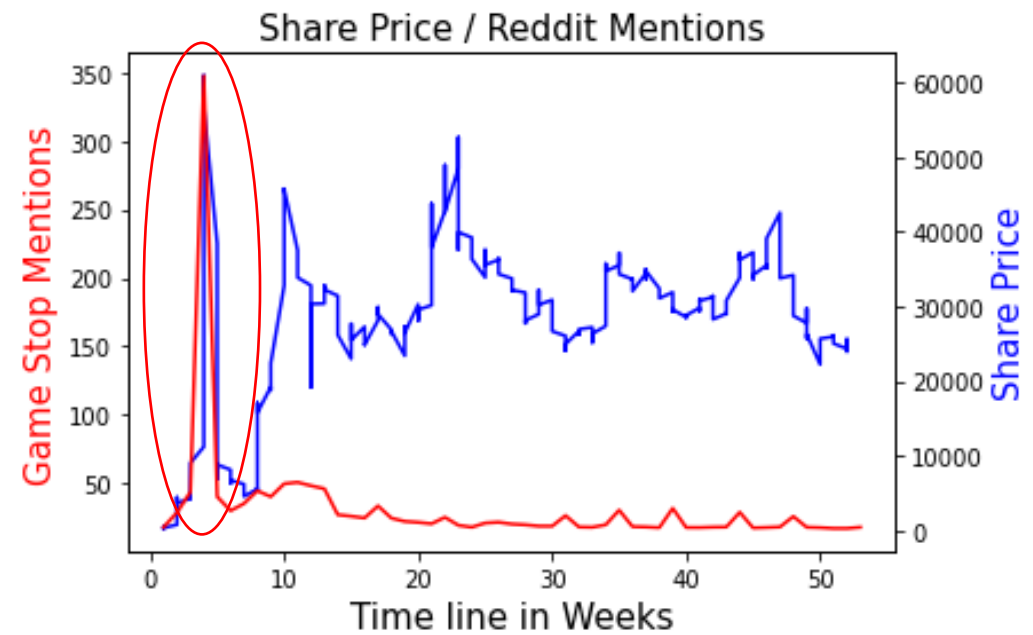
Date Ranges

Rolling Windows

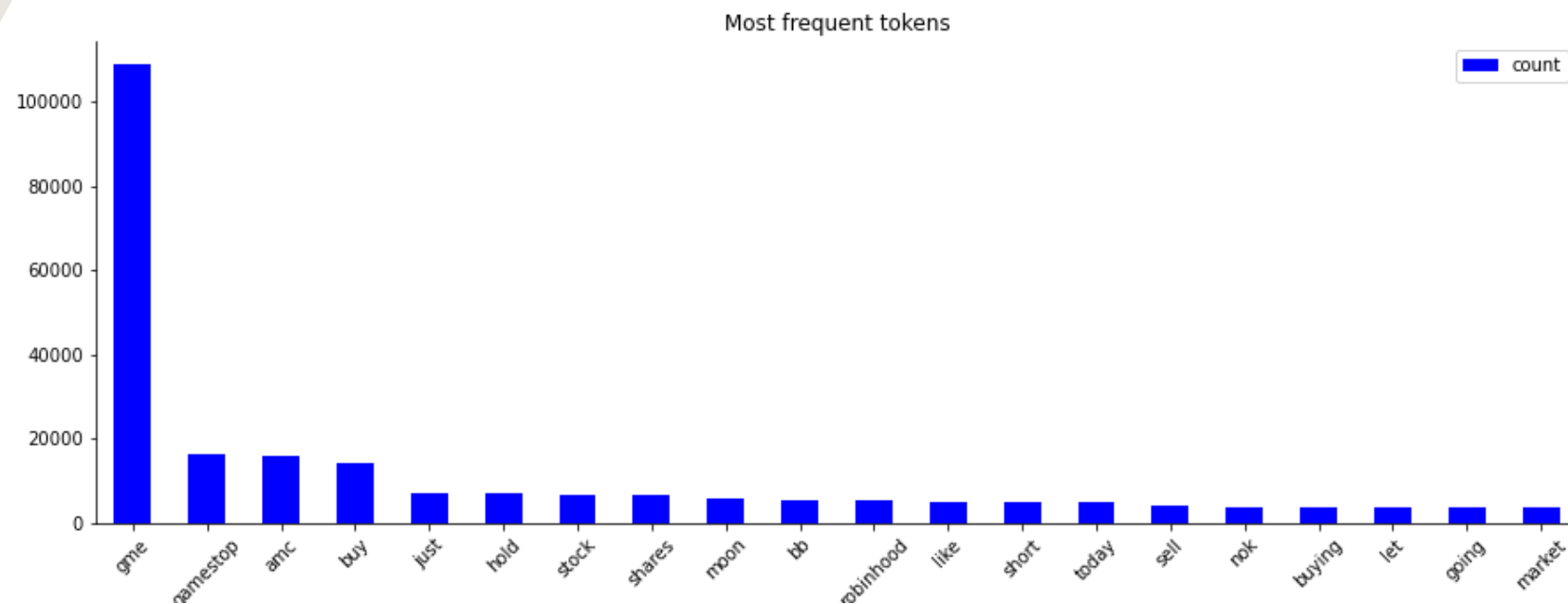
Game Stop Counts



EXPLORATORY DATA ANALYSIS



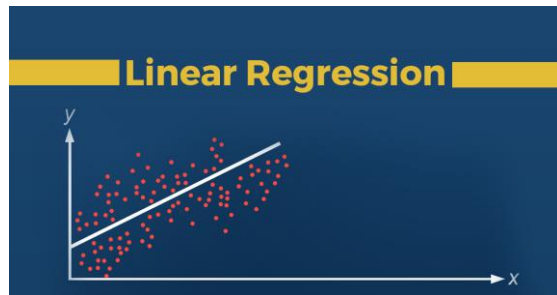
NLP: BAG OF WORDS



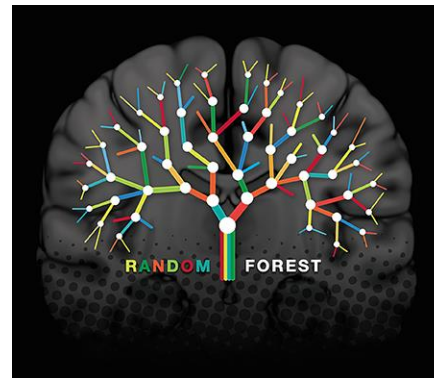
MACHINE LEARNING MODELS

LINEAR REGRESSION

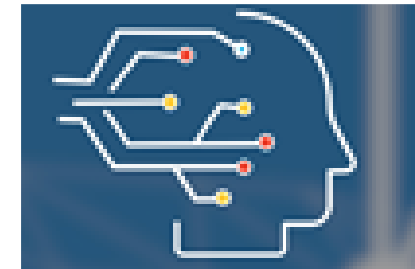
- RIDGE REGRESSION
- LASSO REGRESSION

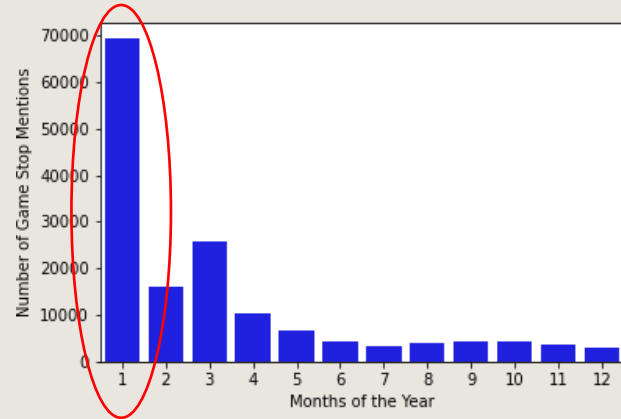


RANDOM FORESTS



XG BOOST





TRAIN / TEST DATA

A – 10 MONTHS / 2 MONTHS

Train 95% - Test 5%

B – 8 MONTHS / 2 MONTHS

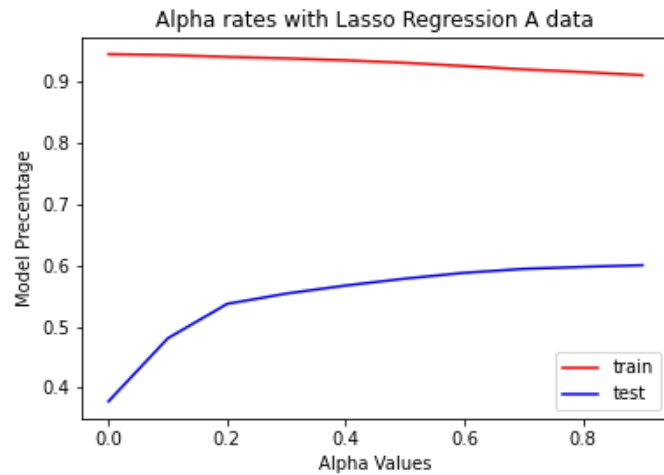
Train 91% - Test 9%

C – 6 MONTHS / 6 MONTHS

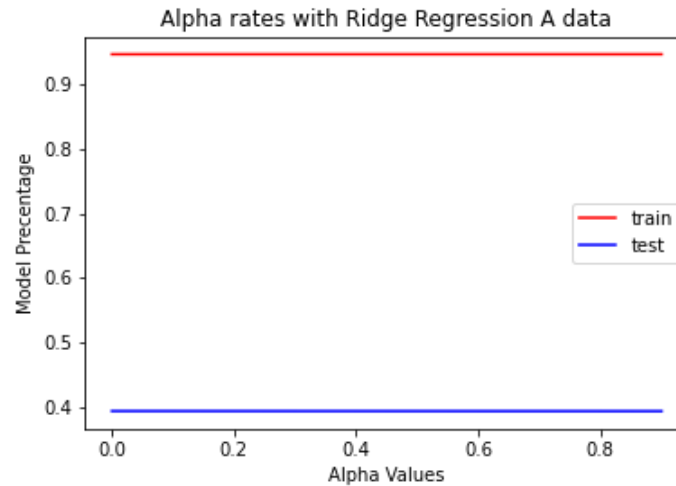
Train 80% - Test 20%

EVALUATION: ACCURACY

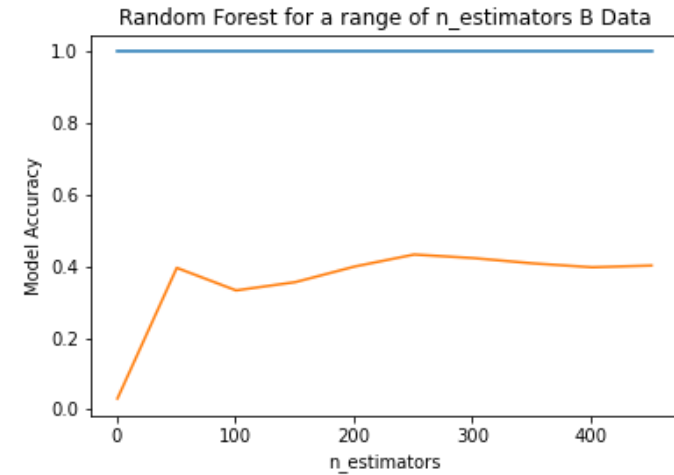
Lasso Regression



Ridge Regression



Random Forest

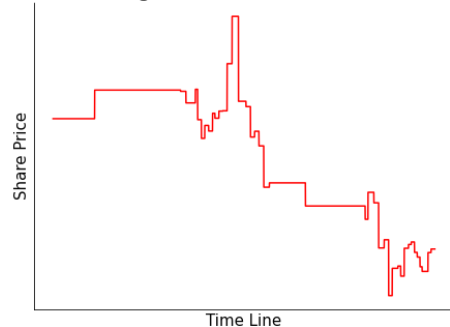


ACCURACY: MODEL SCORES

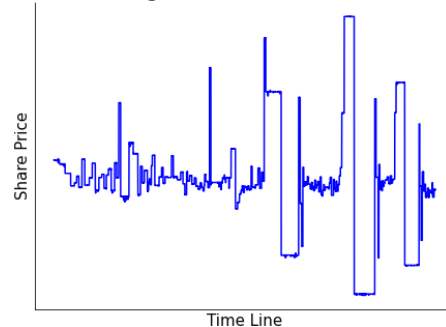
ML Models	Accuracy		ML Models	Accuracy
A_Linear_Regression	0.395258		A_Random_Forest	0.50109
B_Linear_Regression	-0.218658		B_Random_Forest	0.403605
C_Linear_Regression	-10.303093		C_Random_Forest	-0.059489
A_Ridge_Regression	0.395088			
B_Ridge_Regression	-0.217177		A_XG_Boost	0.396404
C_Ridge_Regression	-10.345777		B_XG_Boost	0.340001
A_Lasso_Regression	0.597171		C_XG_Boost	-0.077485
B_Lasso_Regression	0.436013			
C_Lasso_Regression	-2.170383			
A_Lasso_Regression_Alpha_0_9	0.600184			

ACTUAL VALUE VS PREDICTED VALUES

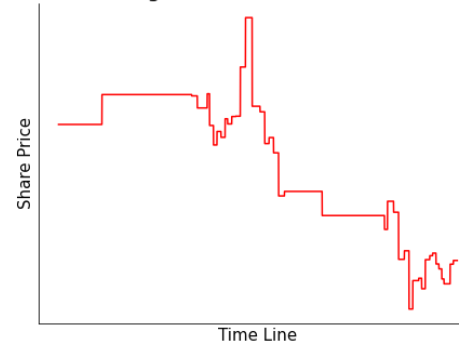
Linear Regression Actual Values C Data



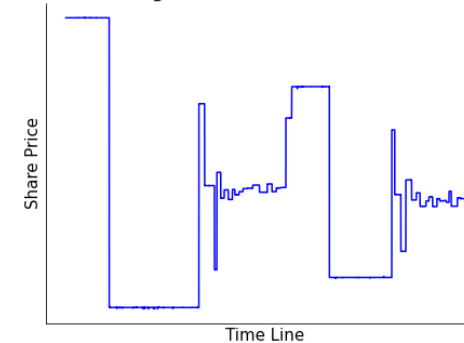
Linear Regression Predict Values Data



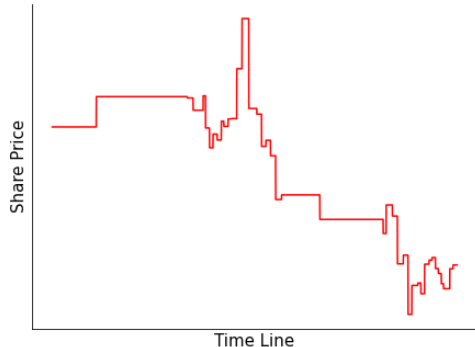
Linear Regression Actual Values A Data



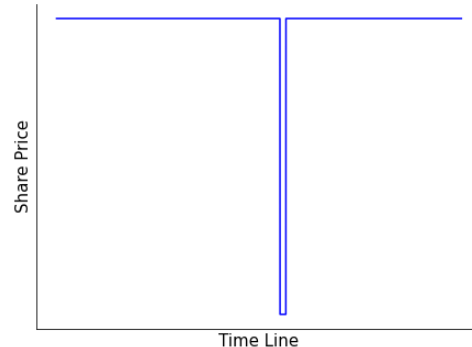
Linear Regression Predict Values A Data



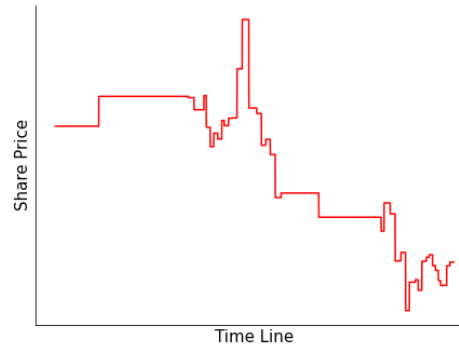
A XG Boost Actual Values C Data



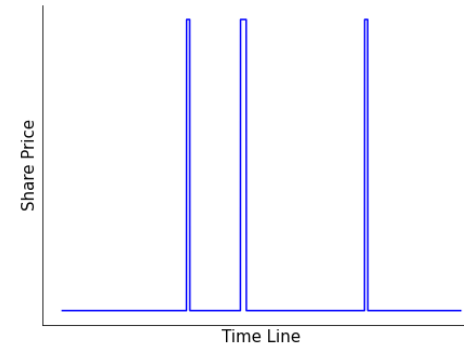
A XG Boost Predict Values Data



B XG Boost Actual Values C Data



B XG Boost Predict Values Data





NEXT STEPS

NLP: SENTIMENT ANALYSIS

Positive and Negative comments

YAHOO FINANCE API

Additional factors that could affect the share price

LARGER SHARE PRICE RANGE

2021 is a peak year



CORRELATION IS NOT CAUSATION

.....until you get high model results.

Thank you!