

Reddit API & NLP

DSI - 13 Project_3

Julian Broche - East Coast cohort

I'm a moderator for a popular Finance related subreddit....I'm lazy, so I built a script/model to detect non-related posts and automatically remove it from my precious forum.

Problem Statement

**Could I build an NLP model
that reduces tinfoil posts on
my assigned subreddit with
high degree of accuracy?**

Let's find out.....

Methodology

- **PRAW** - *Data Sourcing*
- **NLP** - *Text to Vectors*
 - Count Vec
 - TFIDF Vec
- **Models** - *Classification*
 - Logistic Regression
 - Random Forest
 - SVM



Quantitative Findings

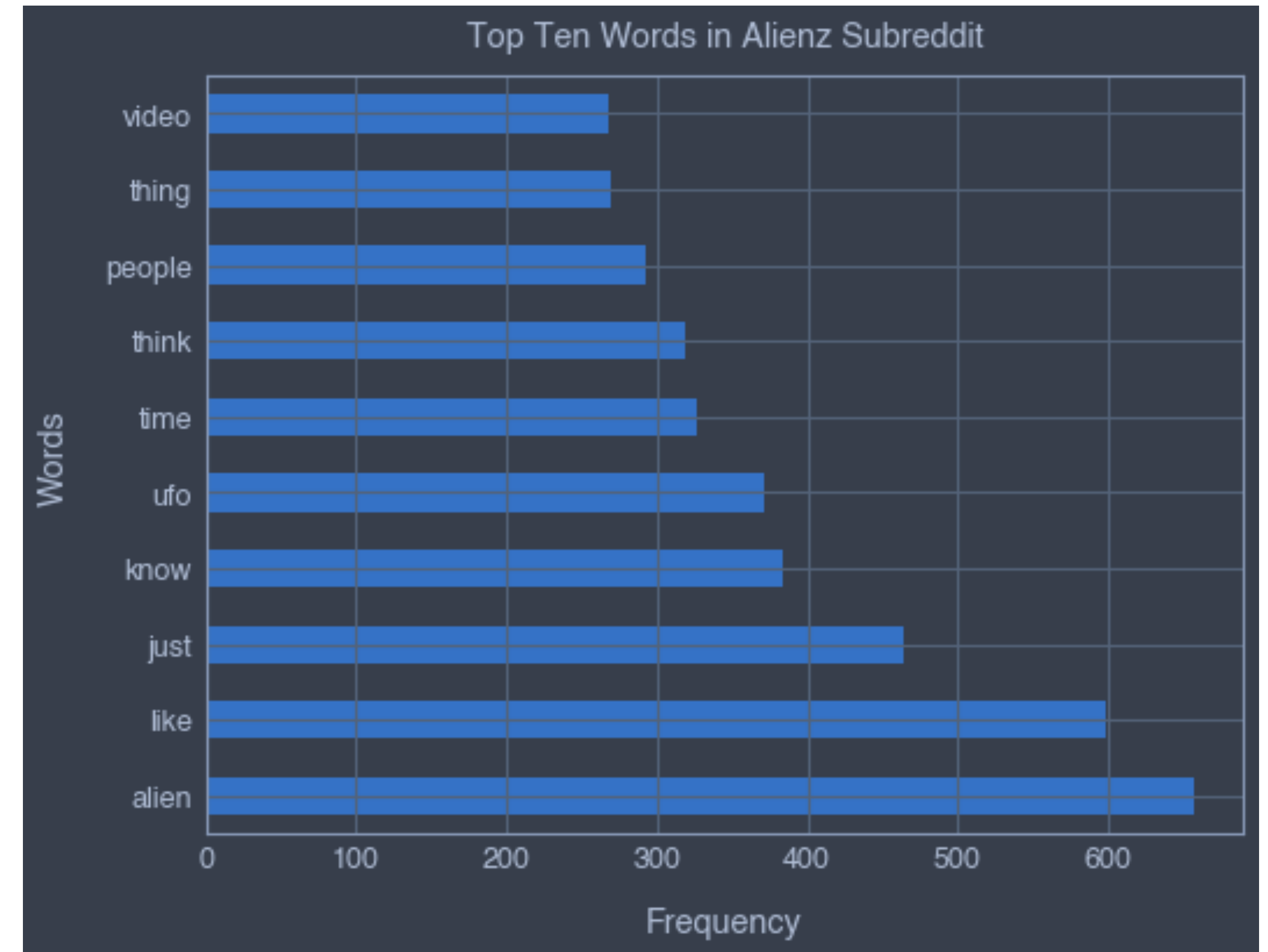
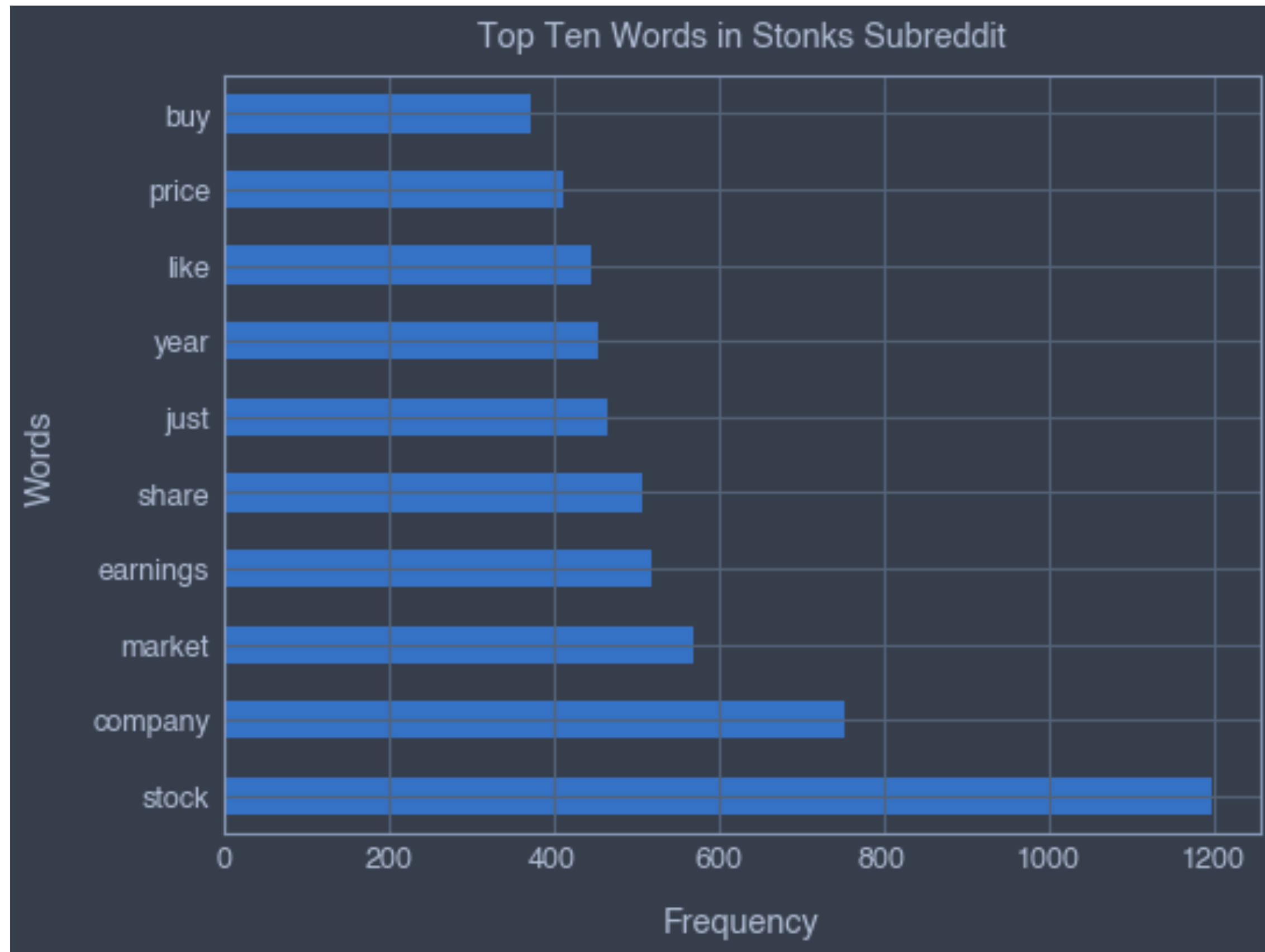
WORDS in Stonks



Words in Alienz



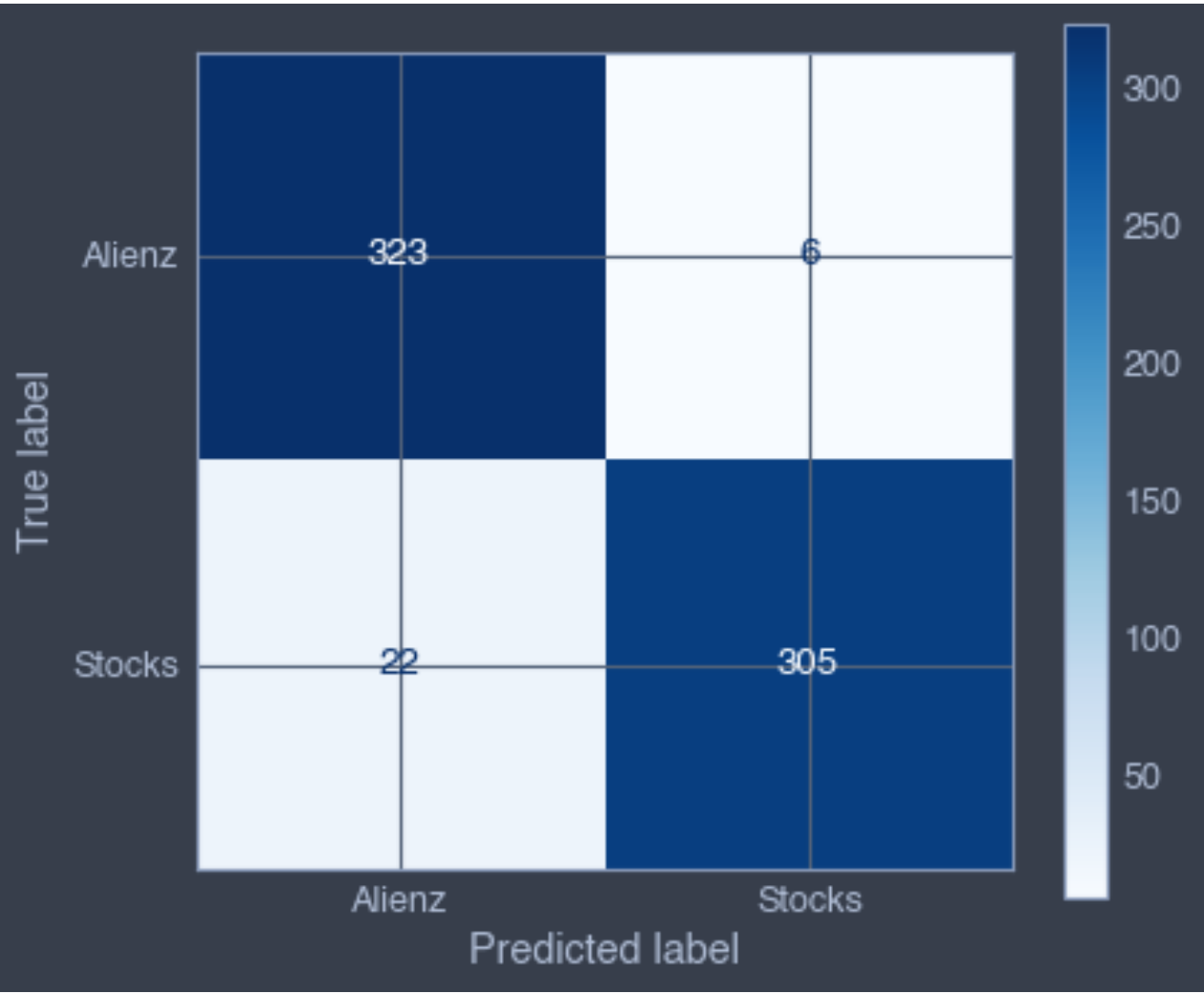
Top 10 words by SubReddit



Lets Talk Models

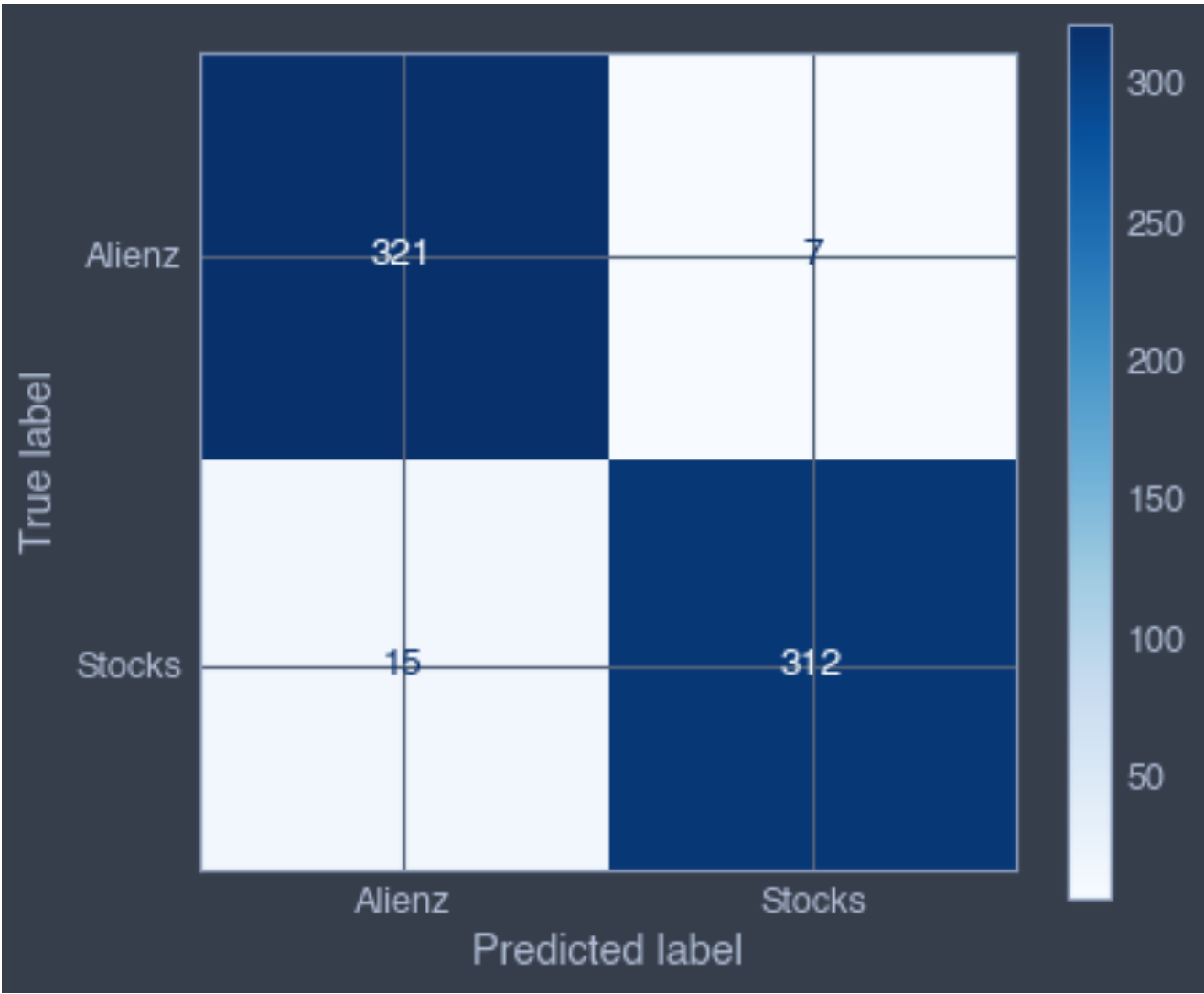
Time to get Confused

Log Reg Matrix



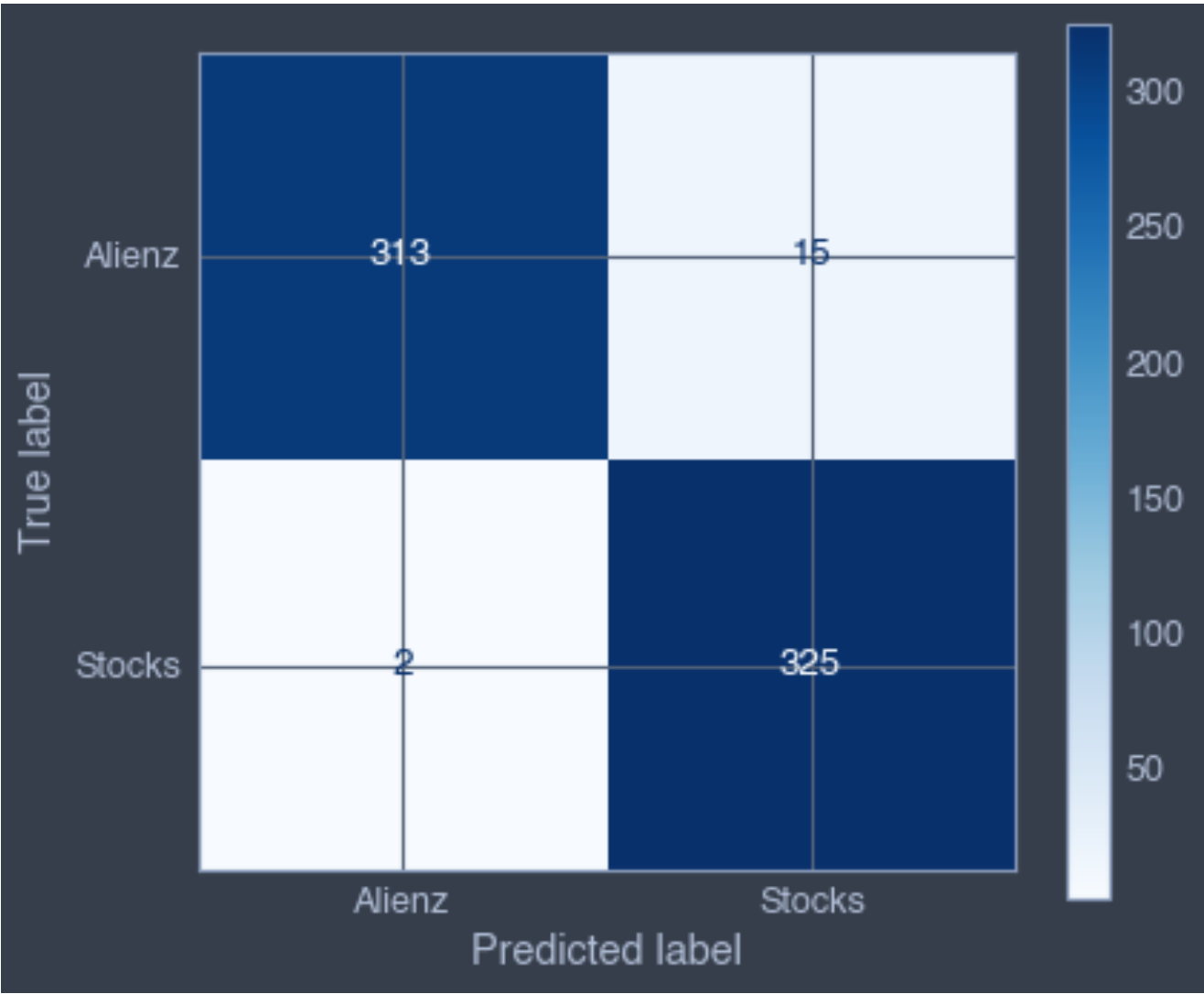
Accuracy Score: 0.957%

Random Forest Matrix



Accuracy Score: 0.966%

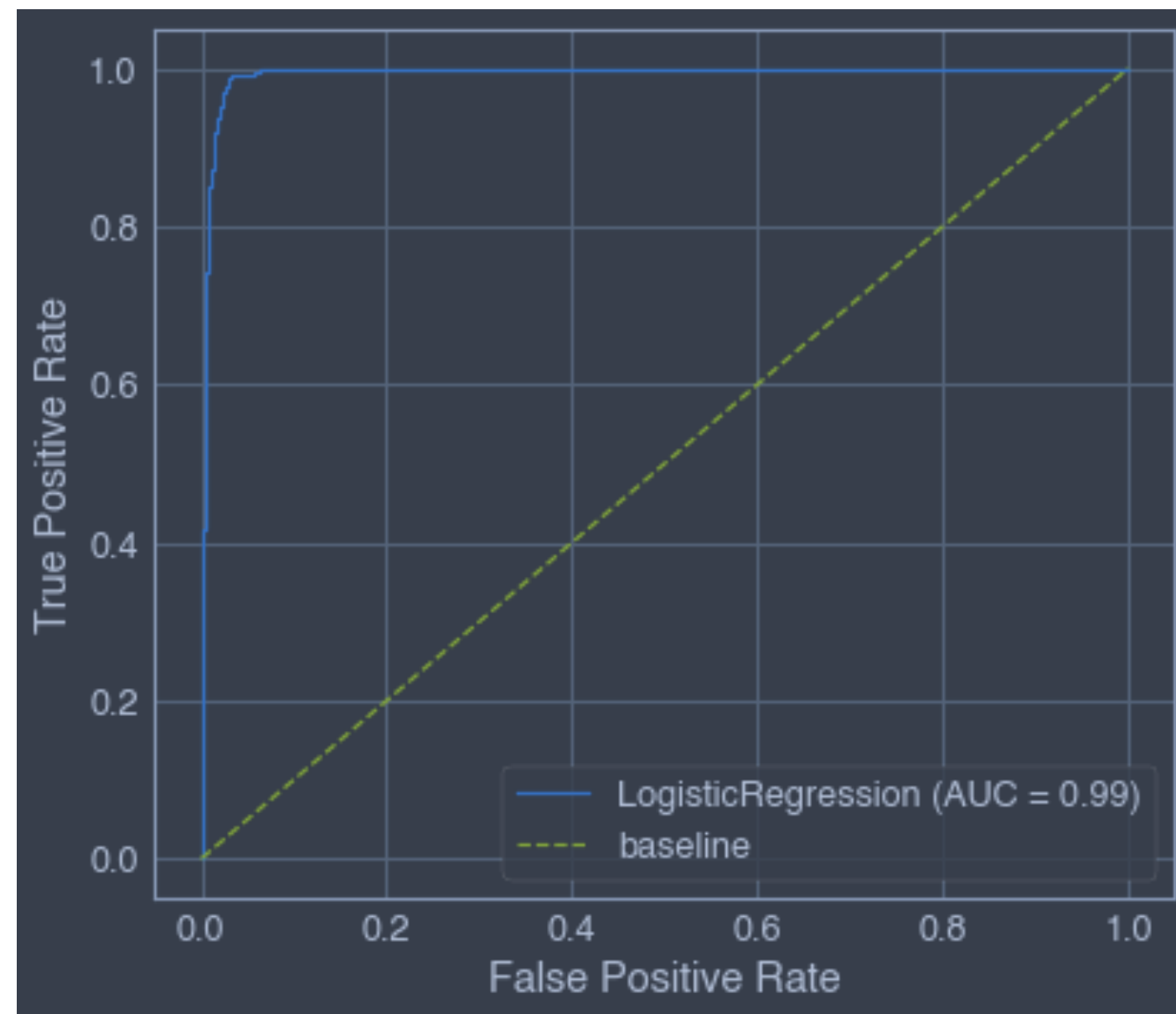
SVM Matrix



Accuracy Score: 0.974%

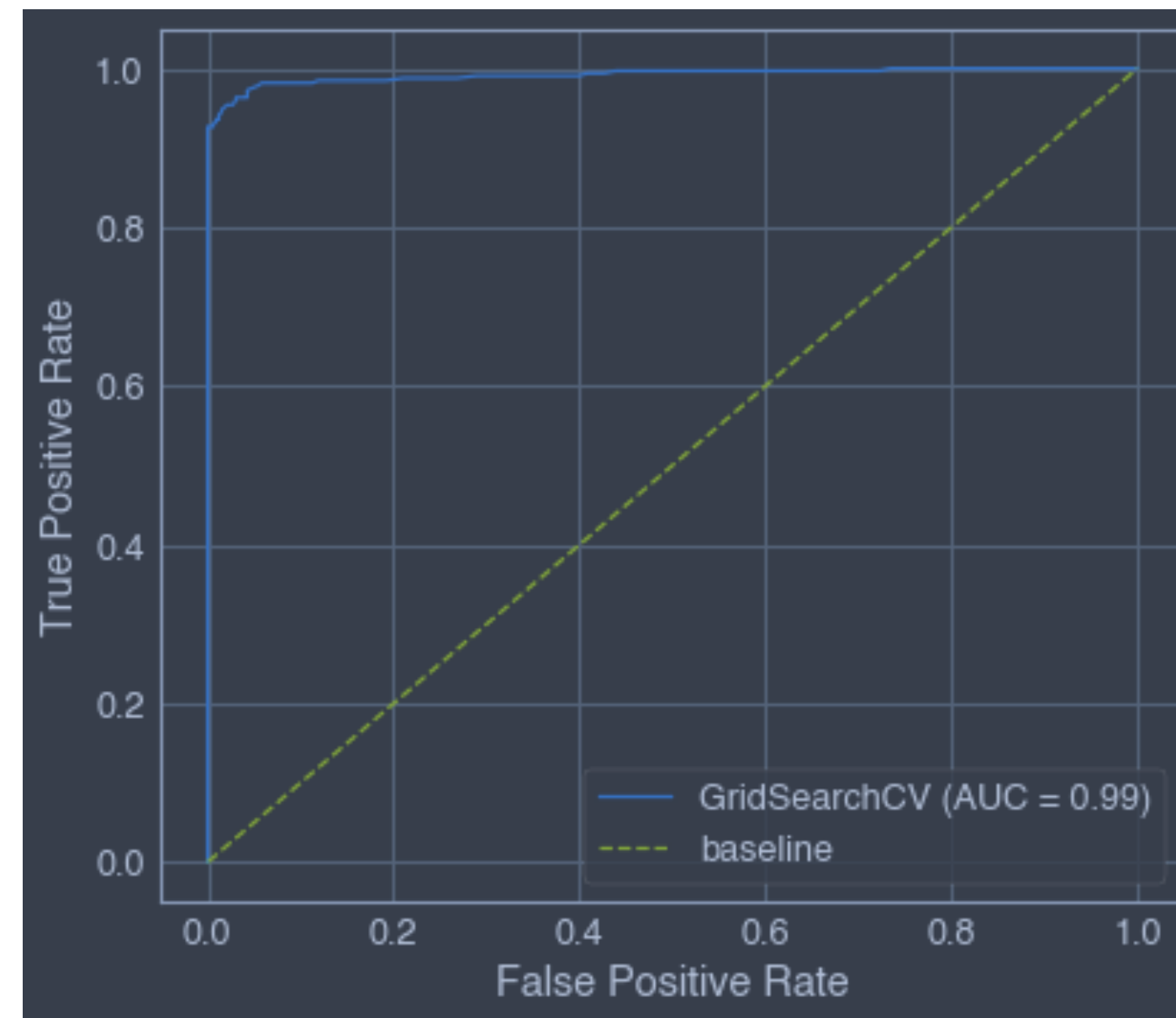
Curves

Log Reg



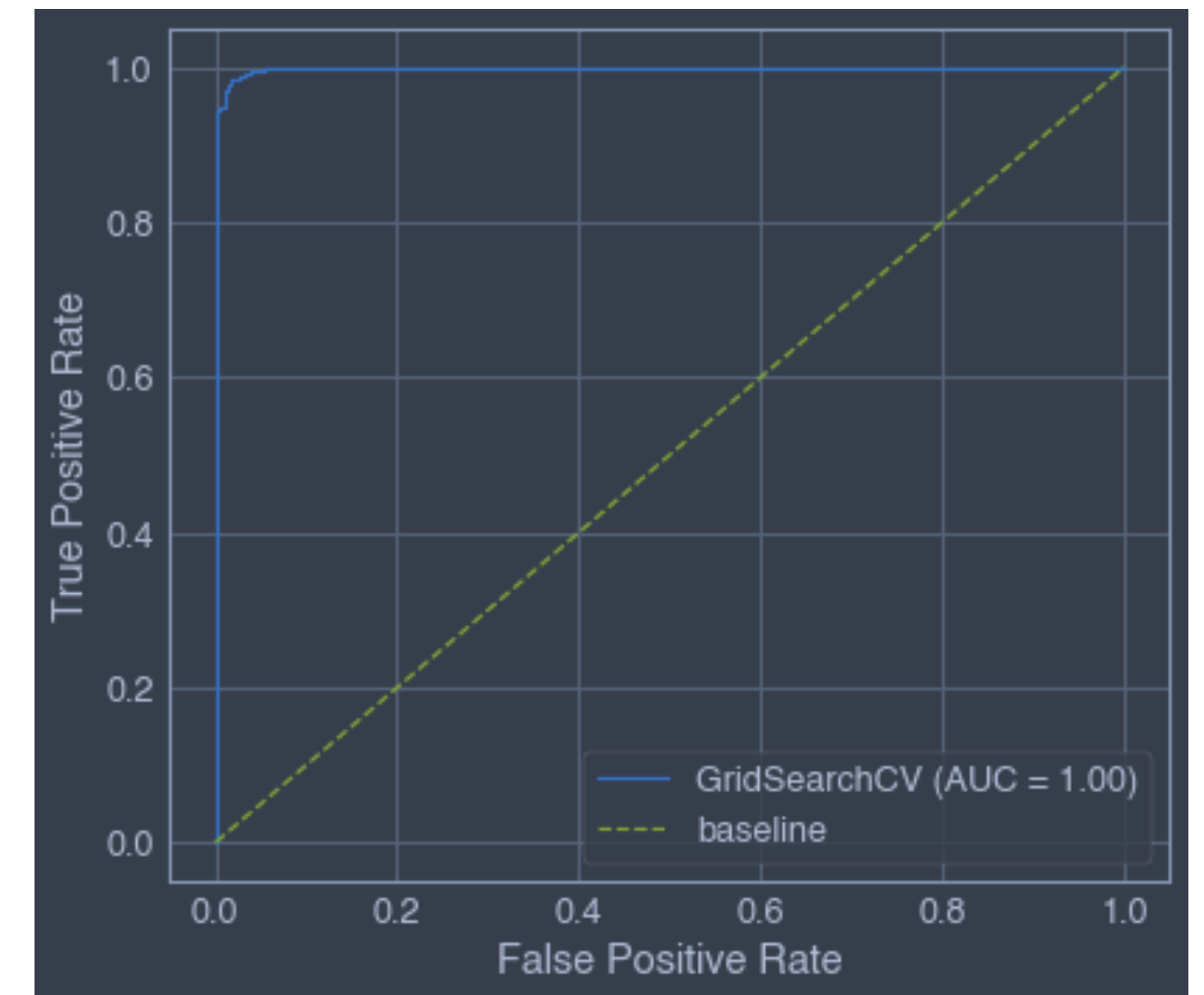
ROC_AUC Score: 0.994%

Random Forest



ROC_AUC Score: 0.991%

SVM



ROC_AUC Score: 0.998%

LogReg vs RF Metrics

Log Reg

Accuracy score 0.9573170731707317					

	Actual Alienz	Actual Stonks			
Predicted Alienz	323	6			
Predicted Stonks	22	305			

	precision	recall	f1-score	support	
0	0.94	0.98	0.96	329	
1	0.98	0.93	0.96	327	
accuracy			0.96	656	
macro avg	0.96	0.96	0.96	656	
weighted avg	0.96	0.96	0.96	656	

ROC_AUC SCORE 0.9948133069351105					

Random Forest

Accuracy score 0.966412213740458					

	Actual Alienz	Actual Stonks			
Predicted Alienz	321	7			
Predicted Stonks	15	312			

	precision	recall	f1-score	support	
0	0.96	0.98	0.97	328	
1	0.98	0.95	0.97	327	
accuracy			0.97	655	
macro avg	0.97	0.97	0.97	655	
weighted avg	0.97	0.97	0.97	655	

ROC_AUC SCORE 0.9919118743939733					

Lets Talk SVM Model

Grid_Best Score: 0.982%
Grid_Best Params: {'C': 10.0, 'gamma': 0.1, 'kernel': 'rbf'}

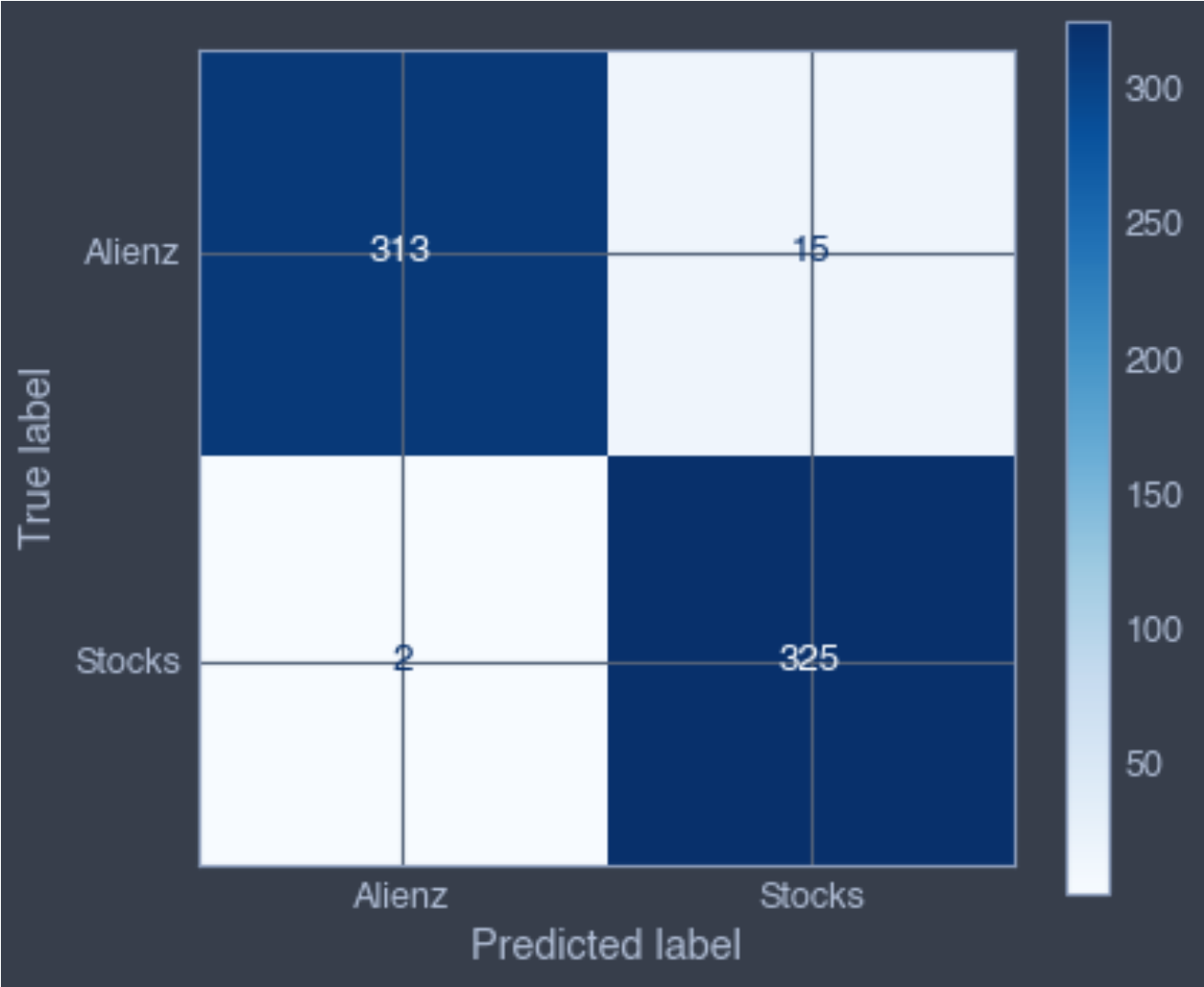
Accuracy score 0.9740458015267176

	Actual Alienz	Actual Stonks
Predicted Alienz	313	15
Predicted Stonks	2	325

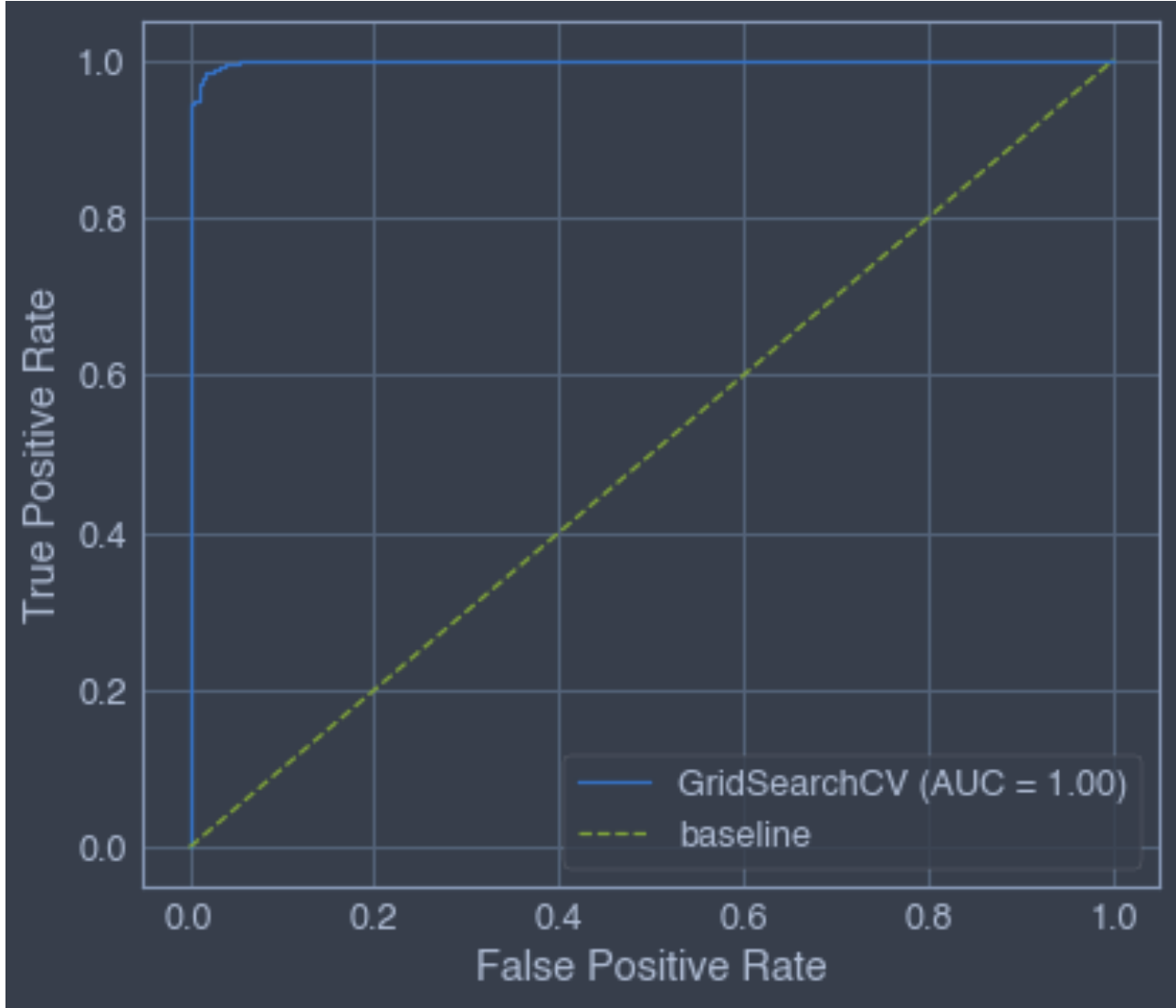
	precision	recall	f1-score	support
0	0.99	0.95	0.97	328
1	0.96	0.99	0.97	327
accuracy			0.97	655
macro avg	0.97	0.97	0.97	655
weighted avg	0.97	0.97	0.97	655

ROC_AUC SCORE 0.9988532110091742

SVM Matrix



ROC_AUC



Recommendation

Don't be posting tin foil talk in the Finance subreddit because it will get swiftly removed.

NOT IN MY HOUSE.

Wrap up Q's

#p3done

Sources

<https://www.reddit.com/r/aliens/>

<https://www.reddit.com/r/stocks/>

Data Visualization - Word Art by Tableau - Julian Broche

***Metrics Print Out By Heather Robbins
Lesson - NLP***