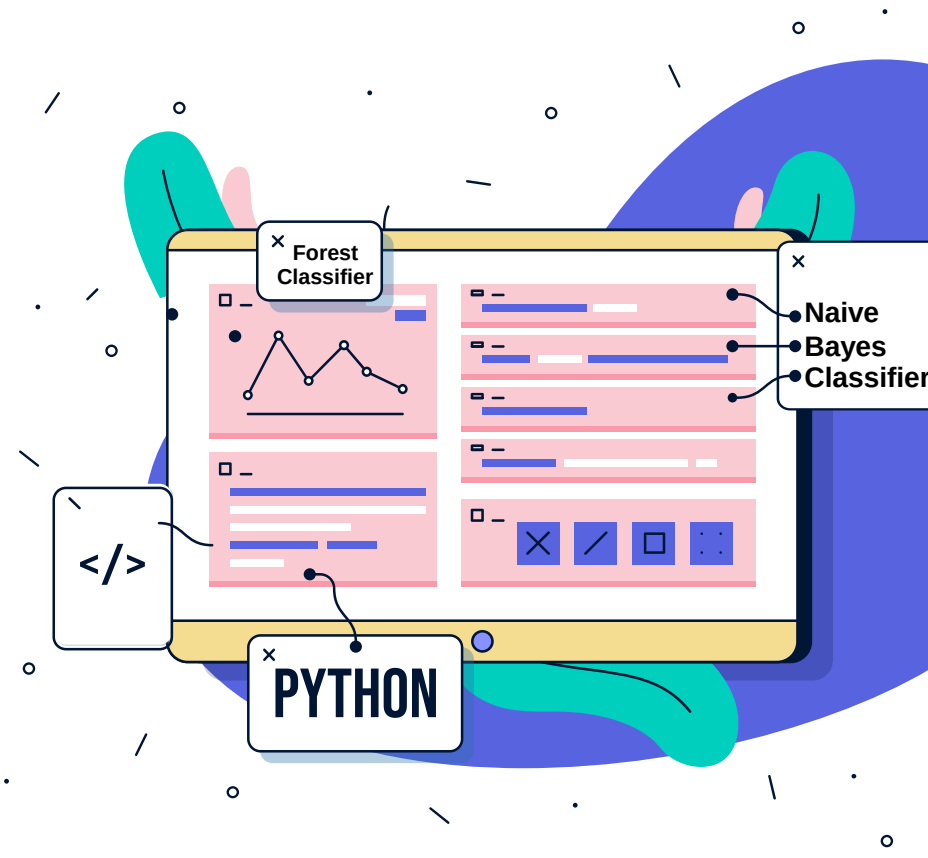
STOCK PRICE PREDICTION

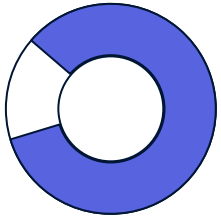
Machine Learning Project



ABSTRACT

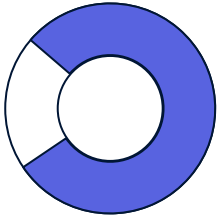
NAIVE BAYES

85% score on the testing set



RANDOM FOREST CLASSIFIER

84% score on the testing set



INTRODUCTION

Predicting whether the stock price is going to increase or not depending on the top 25 headlines of a newspaper.

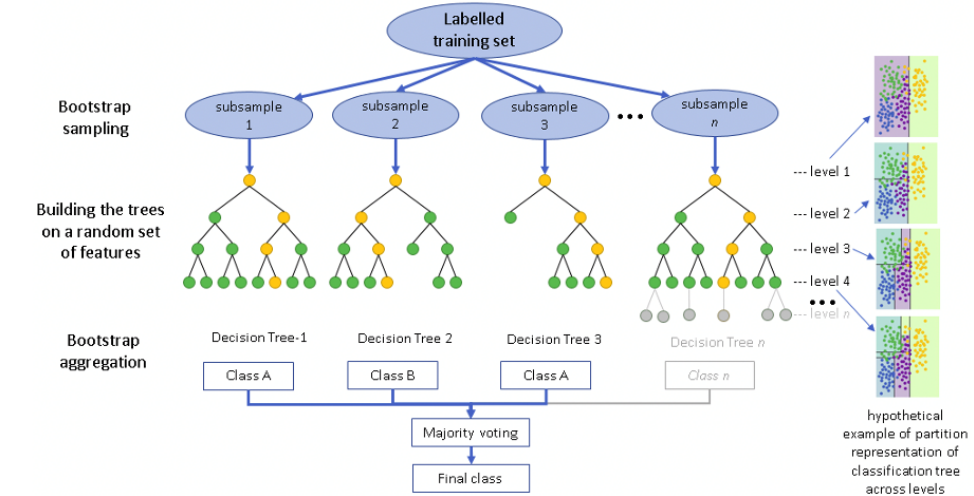
- Trained ML models were able to predict either label 0 or label 1.
- Target value of 0 indicates that the stock price was not going to increase the next day and target value of 1 indicates that the stock price was expected to increase the following day.

RELATED WORK

Four professors from University of California developed a model for stock market forecasting using News Headlines.

They worked on the same dataset as well.

Their main step was to pre-process daily news, to generate features for the prediction problem.



RANDOM FOREST CLASSIFIER

DATASET

NEWS DATA

STOCKS DATA

	Date	Label	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	...	Top16	Top17	To
0	2008-08-08	0	b"Georgia 'downs two Russian warplanes' as cou...	b'BREAKING: Musharraf to be impeached.'	b'Russia Today: Columns of troops roll into So...	b'Russian tanks are moving towards the capital...	b"Afghan children raped with 'impunity,' U.N. ...	b'150 Russian tanks have entered South Ossetia...	b"Breaking: Georgia invades South Ossetia, Rus...	b"The 'enemy combatent' trials are nothing but...	...	b'Georgia Invades South Ossetia - if Russia ge...	b'Al-Qaeda Faces Islamist Backlash'	b'Condolee Rice: "The would no to
1	2008-08-11	1	b'Why wont America and Nato help us? If they w...	b'Bush puts foot down on Georgian conflict'	b"Jewish Georgian minister: Thanks to Israeli ...	b'Georgian army flees in disarray as Russians ...	b"Olympic opening ceremony fireworks 'faked'"	b'What were the Mossad with fraudulent New Zea...	b'Russia angered by Israeli military sale to G...	b'An American citizen living in S.Ossetia blam...	...	b'Israel and the US behind the Georgian aggres...	b'"Do not believe TV, neither Russian nor Geor...	b'Riots are going c Mont (Canad
2	2008-08-12	0	b'Remember that adorable 9-year-old who sang a...	b"Russia 'ends Georgia operation'"	b'"If we had no sexual harassment we would hav...	b"Al-Qa'eda is losing support in Iraq because ...	b'Ceasefire in Georgia: Putin Outmaneuvers the...	b'Why Microsoft and Intel tried to kill the XO...	b'Stratfor: The Russo-Georgian War and the Bal...	b"I'm Trying to Get a Sense of This Whole Geor...	...	b'U.S. troops still in Georgia (did you know t...	b'Why Russias response to Georgia was right'	b'Gorbac accuses of maki "seriou
3	2008-08-13	0	b' U.S. refuses Israel weapons to attack Iran:...	b"When the president ordered to attack Tskhinv...	b' Israel clears troops who killed Reuters cam...	b'Britain\'s policy of being tough on drugs is...	b'Body of 14 year old found in trunk; Latest (...	b'China has moved 10 *million* quake survivors...	b"Bush announces Operation Get All Up In Russi...	b'Russian forces sink Georgian ships '	...	b'Elephants extinct by 2020?'	b'US humanitarian missions soon in Georgia - i...	b"Georg DDOS c from sour

Snippet of the dataset

RESULTS AND DISCUSSION

RANDOM FOREST CLASSIFIER

With a random forest classifier, we got a score of 100% on the training and 85% on the testing set.

```
#Training and testing accuracy
print("Training Accuracy:",accuracy_score(df_train['Label'], y_train_pred)*100)
print("Testing Accuracy:",accuracy_score(df_test['Label'], y_test_pred)*100)
```

```
Training Accuracy: 100.0
Testing Accuracy: 84.92063492063492
```

RESULTS AND DISCUSSION

Classification Report:

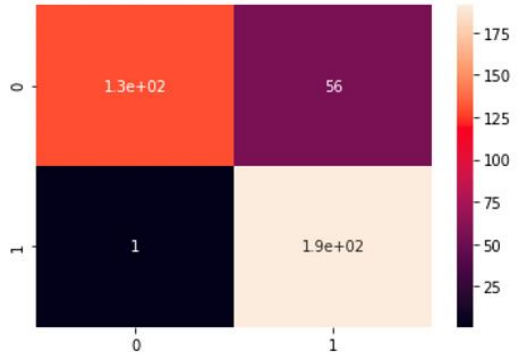
0.85 score for weighted precision, recall and f1 score

```
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.70      0.82       186
           1       0.77      0.99      0.87       192

    accuracy                           0.85       378
   macro avg       0.88      0.85      0.85       378
weighted avg       0.88      0.85      0.85       378
```

Confusion Matrix



RESULTS AND DISCUSSION

NAIVE BAYES CLASSIFIER

With a naive bayes classifier, we got a score of 100% on the training and 84% on the testing set.

```
Training Accuracy: 100.0
Testing Accuracy: 84.12698412698413
```

RESULTS AND DISCUSSION

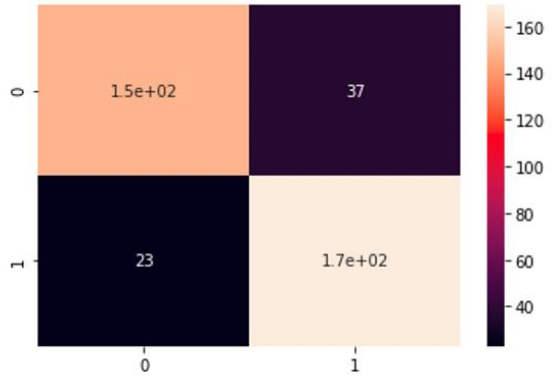
Classification Report:

0.84 score for precision, recall and f1 score

Classification Report:

	precision	recall	f1-score	support
0	0.87	0.80	0.83	186
1	0.82	0.88	0.85	192
accuracy			0.84	378
macro avg	0.84	0.84	0.84	378
weighted avg	0.84	0.84	0.84	378

Confusion Matrix



CONCLUSION

FUTURE WORK

</>

CODE EXPLANATION

THANKS!

Do you have any questions?

Presentation created by Amina Mehić